Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution

CHD Exome Consortium

2019-03

CHD Exome Consortium , Cohorts For Heart Aging Res Genomi , EPIC-CVD Consortium , ExomeBP Consortium , Global Lipids Genetic Consortium , GoT2D Genes Consortium , InterAct , ReproGen Consortium , T2G Genes Consortium , MAGIC Investigators , Justice , A E , Karaderi , T , Highland , H M , Jousilahti , P , Lindström , J , Männistö , S , Perola , M & Tuomilehto , J 2019 , ' Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution ' , Nature Genetics , vol. 51 , no. 3 , pp. 452-+ . https://doi.org/10.1038/s41588-018-0334-2

http://hdl.handle.net/10138/313332
https://doi.org/10.1038/s41588-018-0334-2

unspecified
publishedVersion

ARTICLES
https://doi.org/10.1038/s41588-018-0334-2
nature genetics

Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution

Body-fat distribution is a risk factor for adverse cardiovascular health consequences. We analyzed the association of body-fat distribution, assessed by waist-to-hip ratio adjusted for body mass index, with 228,985 predicted coding and splice site variants available on exome arrays in up to 344,369 individuals from five major ancestries (discovery) and 132,177 European-ancestry individuals (validation). We identified 15 common (minor allele frequency, MAF ≥5%) and nine low-frequency or rare (MAF <5%) coding novel variants. Pathway/gene set enrichment analyses identified lipid particle, adiponectin, abnormal white adipose tissue physiology and bone development and morphology as important contributors to fat distribution, while cross-trait associations highlight cardiometabolic traits. In functional follow-up analyses, specifically in *Drosophila* RNAi-knockdowns, we observed a significant increase in the total body triglyceride levels for two genes (*DNAH10* and *PLXND1*). We implicate novel genes in fat distribution, stressing the importance of interrogating low-frequency and protein-coding variants.

Central body-fat distribution, as assessed by waist-to-hip ratio (WHR), is a heritable and a well-established risk factor for adverse metabolic outcomes[1–6]. Lower values of WHR are associated with lower risk of cardiometabolic diseases such as type 2 diabetes[7,8], or differences in bone structure and gluteal muscle mass[9]. These epidemiological associations are consistent with our previously reported genome-wide association study (GWAS) results of 49 loci associated with WHR (after adjusting for body mass index, WHRadjBMI)[10]. Notably, genetic predisposition to higher WHRadjBMI is associated with increased risk of type 2 diabetes and coronary heart disease (CHD), which seems to be causal[9].

Recently, large-scale studies have identified roughly 125 common loci for multiple measures of central obesity, primarily non-coding variants of relatively modest effect[10–16]. Large-scale interrogation of coding and splice site variants, including both common (minor allele frequency, MAF ≥5%) and low-frequency or rare variants (LF/RV: MAF <5%), may lead to additional insights into the etiology of central obesity. Here, we identify and characterize such variants associated with WHRadjBMI using ExomeChip array genotypes.

Results

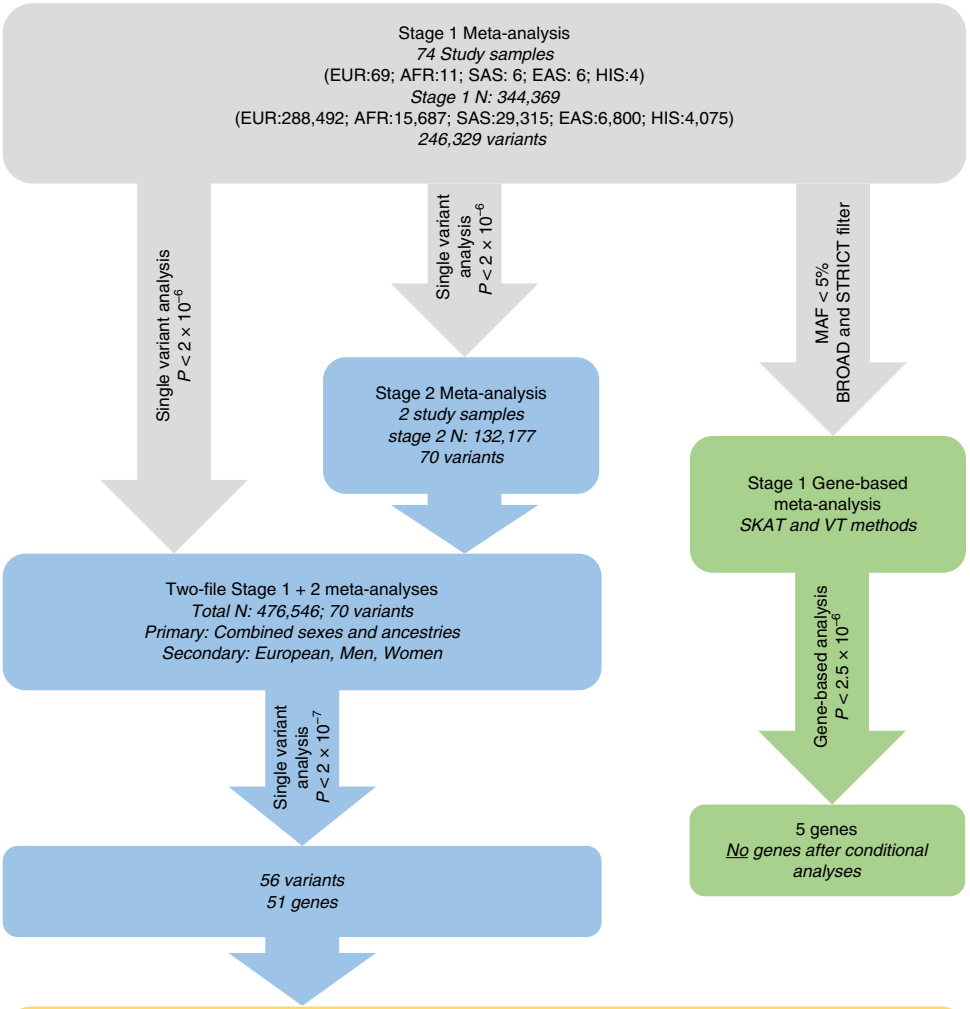
Protein-coding and splice site variation associations. We conducted a two-stage fixed-effects meta-analysis testing additive and recessive models to detect protein-coding genetic variants that influence WHRadjBMI (Methods and Fig. 1). Stage 1 included up to 228,985 variants (218,195 low-frequency or rare) in up to 344,369 individuals from 74 studies of European, South and East Asian, African and Hispanic/Latino descent individuals (Supplementary Data 1–3). Stage 2 assessed 70 suggestive ($P<2\times10^{-6}$) stage 1 variants in two cohorts, UK Biobank (UKBB) and deCODE for a total stage 1 and 2 sample size of 476,546 (88% European). Of the 70 variants considered, two common and five low-frequency or rare were not available in stage 2 (Tables 1 and 2 and Supplementary Data 4–6). Variants are considered novel and statistically significant if they were greater than one megabase (Mb) from a previously identified WHRadjBMI SNP[10–16] and achieve array-wide significance ($P<2\times10^{-7}$, stage 1 and 2).

In our primary meta-analysis, including all stage 1 and 2 samples, we identified 48 coding variants (16 novel) across 43 genes, 47 assuming an additive model and one under a recessive model (Table 1 and Supplementary Figs. 1–4). Due to possible heterogeneity, we also performed European-only meta-analysis. Here, four additional coding variants were significant (three novel) assuming an additive model (Table 1 and Supplementary Figs. 5–8). Of these 52 significant variants, 11 were low-frequency or rare and displayed larger effect estimates than many previously reported common variants[10], including seven novel variants in *RAPGEF3, FGFR2, R3HDML, HIST1H1T, PCNXL3, ACVR1C* and *DARS2*. Variants with MAF ≤1% had effect sizes approximately three times greater than those of common variants (MAF ≥5%). Despite large sample size, we cannot rule out the possibility that additional low-frequency or rare variants with smaller effects exist (see estimated 80% power in Fig. 2). However, in the absence of common variants with similarly large effects, our results point to the importance of investigating LF/RVs.

Given established sex differences in the genetic underpinnings of WHRadjBMI[10,11], we also performed sex-stratified analyses. We detected four additional novel variants that exhibit significant sex-specific effects ($P_{sexhet}<7.14\times10^{-4}$, Methods) in *UGGT2* and *MMP14* for men; and *DSTYK* and *ANGPTL4* for women (Table 2 and Supplementary Figs. 9–15); including low-frequency or rare in *UGGT2* and *ANGPTL4* ($MAF_{men}=0.6\%$ and $MAF_{women}=1.9\%$, respectively). Additionally, 14 variants from the sex-combined meta-analyses displayed significantly stronger effects in women, including the novel, low-frequency *ACVR1C* variant (rs55920843, MAF = 1.1%). Overall, 19 of the 56 variants (32%) identified across all meta-analyses (48 from all ancestry, 4 from European-only and 4 from sex-stratified analyses) showed significant sex-specific effects on WHRadjBMI: 16 variants with significantly stronger effects in women and three in men (Fig. 1).

In summary, we identified 56 array-wide significant coding variants ($P<2.0\times10^{-7}$); 43 common (14 novel) and 13 low-frequency or rare (9 novel). For the 55 significant variants from the additive model, we examined potential collider bias[17,18] (Methods, Supplementary Table 1 and Supplementary Note). Overall, 51 of 55 variants were robust to collider bias[17,18]. Of these, 25 variants were nominally associated with BMI ($P_{BMI}<0.05$), yet effect sizes changed little after correction for potential biases (15% change in

A full list of authors and affiliations appears at the end of the paper.



Number of significant variants (all ancestries/european)

Strata	ADD	REC	NOVEL*	GWAS INDEP¥
Combined	47/4	1/0	16/3	1/0
Women	16/0	0/0	3/0	0/0
Men	2/0	0/0	2/0	0/0
Total non-overlapping signals	56 variants across 41 signals		24 new signals	

*Novel variants include those that are >1 Mb from a previously published WHRadjBMI GWAS tag SNP.
¥ Independent (INDEP) includes variants that are nearby known WHRadjBMI GWAS tag variants, but were determined independent after conditional analysis.

Fig. 1 | Summary of meta-analysis study design and workflow. Abbreviations: EUR, European, AFR, African, SAS, South Asian, EAS, East Asian and HIS, Hispanic/Latino ancestry.

effect estimate on average). For four of the 55 SNPs (rs141845046, rs1034405, rs3617, rs9469913), attenuation following correction was noted ($P_{corrected} > 9 \times 10^{-4}$, 0.05/55), including one novel variant, rs1034405 in *C3orf18*, demonstrating a possible overestimation of these effects in the current analysis.

Using stage 1 results, we then aggregated LF/RVs across genes and tested their joint effect with SKAT and burden tests[19] (Supplementary Table 2 and Methods). None of the five genes that reached array-wide significance ($P < 2.5 \times 10^{-6}$, 0.05/16,222 genes tested: *RAPGEF3*, *ACVR1C*, *ANGPTL4*, *DNAI1* and *NOP2*) remained significant after conditioning on the most significant single variant.

Conditional analyses. We next implemented conditional analyses to determine (1) the total number of independent signals identified and (2) whether the 33 variants near known GWAS signals (<+/−1 Mb) represent independent novel associations. We used approximate joint conditional analyses to test for independence in

Table 1 | Association results for combined sexes. Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-7}$) in the sex-combined meta-analyses

Locus (+/− 1 Mb of a given variant)	Chr:position (GRCh37)[b]	rsID	EA	OA	Gene[c]	Amino acid change[c]	If locus is known, nearby (<1 Mb) published variant(s)[d]	*N*	EAF	β[e]	s.e.m.	*P* value	*P* value for sex-heterogeneity[f]	Other criteria for significance[h]
Variants in novel loci														
All ancestry additive model sex-combined analyses														
1	2:158412701	rs55920843	T	G	*ACVR1C*	N150H	–	455,526	0.989	0.065	0.011	**4.8E-10**	**1.7E-07**	
2	3:50597092	rs1034405	G	A	*C3orf18*	A162V	–	455,424	0.135	0.016	0.003	**1.9E-07**	8.8E-01	G,C
3	4:120528327	rs3733526	G	A	*PDE5A*	A41V	–	461,521	0.187	0.015	0.003	**2.6E-08**	5.2E-03	
4	6:26108117	rs146860658	T	C	*HIST1H1T*	A69T	–	217,995	0.001	0.229	0.042	**4.3E-08**	6.3E-01	S
5	7:6449496	rs2303361	C	T	*DAGLB*	Q664R	–	475,748	0.221	0.014	0.003	**6.2E-08**	3.4E-03	G
6	10:123279643	rs138315382	T	C	*FGFR2*	synonymous	–	236,962	0.001	0.258	0.049	**1.4E-07**	1.1E-01	G,S
7	11:65403651	rs7114037	C	A	*PCNXL3*	H1822Q	–	448,861	0.954	0.029	0.005	**1.8E-08**	4.4E-01	
8	12:48143315	rs145878042	A	G	*RAPGEF3*	L300P	–	470,513	0.990	0.085	0.010	**7.2E-17**	7.3E-03	
9	12:108618630	rs3764002	C	T	*WSCD2*	T266I	–	474,637	0.737	0.014	0.002	**9.8E-10**	5.5E-01	
10	15:42032383	rs17677991	G	C	*MGA*	P1523A	–	469,874	0.345	0.015	0.002	**3.5E-11**	9.1E-01	
11	16:4432029	rs3810818	A	C	*VASN*	E384A	–	424,163	0.231	0.016	0.003	**2.0E-09**	3.3E-01	
	16:4445327	rs3747579	C	T	*CORO7*	R193Q	–	453,078	0.299	0.018	0.002	**2.2E-13**	4.3E-02	
	16:4484396	rs1139653	A	T	*DNAJA3*	N75Y	–	434,331	0.284	0.015	0.002	**4.3E-10**	1.4E-01	
12	19:49232226	rs2287922	A	G	*RASIP1*	R601C	–	430,272	0.494	0.014	0.002	**1.6E-09**	3.7E-02	
	19:49244220	rs2307019	G	A	*IZUMO1*	A333V	–	476,147	0.558	0.012	0.002	**4.7E-08**	3.9E-02	
13	20:42965811	rs144098855	T	C	*R3HDML*	P5L	–	428,768	0.001	0.172	0.032	**9.7E-08**	1.0E+00	G
European ancestry additive model sex-combined analyses														
14	1:173802608	rs35515638	G	A	*DARS2*	K196R	–	352,646	0.001	0.201	0.038	**1.4E-07**	6.0E-02	G
15	14:58838668	rs1051860	A	G	*ARID4A*	synonymous	–	367,079	0.411	0.013	0.002	**2.2E-08**	1.3E-01	
16	15:42115747	rs3959569	C	G	*MAPKBP1*	R1240H	–	253,703	0.349	0.017	0.003	**2.0E-08**	6.3E-01	
Variants in previously identified loci														
All ancestry additive model sex-combined analyses														
1	1:119427467	rs61730011	A	C	*TBX15*	M566R	rs2645294, rs12731372, rs12143789, rs1106529	441,461	0.957	0.041	0.005	**2.2E-14**	6.7E-01	
	1:119469188	rs10494217	T	G		H156N		472,259	0.174	0.018	0.003	**1.4E-10**	6.0E-01	
2	1:154987704	rs141845046	C	T	*ZBTB7B*	P190S	rs905938	476,440	0.976	0.037	0.007	**3.8E-08**	**7.9E-07**	C
3	2:165551201	rs7607980	T	C	*COBLL1*	N941D	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	389,883	0.879	0.026	0.004	**1.6E-13**	**3.0E-30**	
4	2:188343497	rs7586970	T	C	*TFPI*	N221S	rs1569135	452,638	0.697	0.016	0.002	**3.0E-12**	6.3E-01	
5	3:52558008	rs13303	T	C	*STAB1*	M113T	rs2276824	470,111	0.445	0.019	0.002	**5.5E-18**	6.7E-02	
	3:52833805	rs3617	C	A	*ITIH3*	Q315K		452,150	0.541	0.015	0.002	**1.6E-12**	4.0E-01	C
6	3:129137188	rs62266958	C	T	*EFCAB12*	R197H	rs10804591	476,382	0.936	0.036	0.004	**8.3E-17**	**9.3E-05**	
	3:129284818	rs2625973	A	C	*PLXND1*	L1412V		476,338	0.733	0.016	0.002	**9.2E-11**	**1.6E-05**	
7	4:89625427	rs1804080	G	C	*HERC3*	E946Q	rs9991328	446,080	0.838	0.021	0.003	**1.5E-12**	**4.1E-06**	
	4:89668859	rs7657817	C	T	*FAM13A*	V443I		476,383	0.815	0.016	0.003	**5.0E-09**	**9.6E-05**	
8	5:176516631	rs1966265	A	G	*FGFR4*	V10I	rs6556301	455,246	0.236	0.023	0.003	**1.7E-19**	2.1E-01	
9	6:7211818	**rs1334576**[g]	G	A	*RREB1*	G195R	rs1294410	451,044	0.565	0.017	0.002	**3.9E-15**	1.5E-01	
10	6:34827085	rs9469913	A	T	*UHRF1BP1*	Q984H	rs1776897	309,684	0.847	0.021	0.004	**1.2E-08**	2.7E-01	C

Continued

Table 1 | Association results for combined sexes. Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-7}$) in the sex-combined meta-analyses (Continued)

Locus (+/− 1 Mb of a given variant)	Chr:position (GRCh37)[b]	rsID	EA	OA	Gene[c]	Amino acid change[c]	If locus is known, nearby (<1 Mb) published variant(s) [d]	*N*	EAF	β[e]	s.e.m.	*P* value	*P* value for sex-heterogeneity[f]	Other criteria for significance[h]
11	6:127476516	rs1892172	A	G	*RSPO3*	synonymous	rs11961815, rs72959041, rs1936805	476,358	0.543	0.031	0.002	**2.6E-47**	**7.7E-09**	
	6:127767954	**rs139745911**[g]	A	G	*KIAA0408*	P504S		391,469	0.010	0.103	0.012	**6.8E-19**	**2.0E-04**	
12	7:73012042	rs35332062	G	A	*MLXIPL*	A358V	rs6976930	451,158	0.880	0.020	0.003	**1.8E-09**	1.5E-01	
	7:73020337	rs3812316	C	G		Q241H		454,738	0.881	0.021	0.003	**2.0E-10**	5.8E-02	
13	10:95931087	rs17417407	T	G	*PLCE1*	R240L	rs10786152	476,475	0.173	0.018	0.003	**2.5E-11**	5.9E-01	
14	11:64031241	rs35169799	T	C	*PLCB3*	S778L	rs11231693	476,457	0.061	0.034	0.004	**9.1E-15**	**1.3E-04**	
15	12:123444507	rs58843120	G	T	*ABDB9*	F92L	rs4765219, rs863750	466,498	0.987	0.053	0.009	**1.3E-08**	3.5E-01	
	12:124265687	rs11057353	T	C	*DNAH10*	S228P		476,360	0.373	0.018	0.002	**2.1E-16**	**2.7E-08**	
	12:124330311	rs34934281	C	T		T1785M		476,395	0.889	0.025	0.003	**2.9E-14**	**3.1E-08**	
	12:124427306	rs11057401	T	A	*CCDC92*	S53C		467,649	0.695	0.029	0.002	**7.3E-37**	**5.5E-11**	
16	15:56756285	rs1715919	G	T	*MNS1*	Q55P	rs8030605	476,274	0.096	0.023	0.004	**8.8E-11**	2.7E-02	
17	16:67397580	rs9922085	G	C	*LRRC36*	R101P	rs6499129	469,474	0.938	0.034	0.005	**3.8E-13**	5.9E-01	
	16:67409180	rs8052655	G	A		G388S		474,035	0.939	0.034	0.005	**5.5E-13**	4.0E-01	
18	19:18285944	rs11554159	A	G	*IFI30*	R76Q	rs12608504	476,389	0.257	0.015	0.002	**3.5E-10**	3.1E-03	
	19:18304700	rs874628	G	A	*MPV17L2*	M72V		476,388	0.271	0.015	0.002	**1.2E-10**	2.5E-03	
19	20:33971914	rs4911494	T	C	*UQCC1*	R51Q	rs224333	451,064	0.602	0.018	0.002	**2.5E-16**	1.5E-03	
	20:34022387	rs224331	A	C	*GDF5*	S276A		345,805	0.644	0.017	0.003	**1.8E-11**	3.2E-03	
All ancestry recessive model sex-combined analyses														
20	17:17425631	rs897453	C	T	*PEMT*	V58L	rs4646404	476,546	0.569	0.025	0.004	**4.1E-11**	8.2E-01	
European ancestry additive model sex-combined analyses														
6	3:129293256	rs2255703	T	C	*PLXND1*	M870V	rs10804591	420,520	0.620	0.014	0.002	**3.1E-09**	**1.6E-04**	

Abbreviations: GRCh37, human genome assembly build 37; rsID, based on dbSNP; VEP, Ensembl Variant Effect Predictor toolset; GTEx, Genotype-Tissue Expression project; s.d., standard deviation; s.e.m., standard error of the mean; *N*, sample size; EAF, effect allele frequency; EA, effect allele; OA, other allele. [a]Coding variants refer to variants located in the exons and splicing junction regions. [b]Variant positions are reported according to human genome assembly build 37 and their alleles are coded based on the positive strand. [c]The gene the variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on the VEP toolset and transcript abundance from GTEx database). [d]Previously published variants within +/−1 Mb are from Shungin et al.[10], except for rs6976930 and rs10786152 from Graff et al.[14] and rs6499129 from Ng. et al.[16]. [e]Effect size is based on the s.d. per effect allele [f]*P* value for sex heterogeneity, testing for difference between women-specific and men-specific beta estimates and standard errors, was calculated using EasyStrata: Winkler et al.[11]. Bold *P* values met significance threshold after Bonferroni correction ($P < 7.14E-04$; that is, 0.05/70 variants). [g]**rs1334576 in *RREB1*** is a new signal in a known locus that is independent from the known signal, rs1294410 (see Supplementary Table 4 and Supplementary Data 7); **rs139745911 in *KIAA0408*** is a secondary signal in a known locus (see Supplementary Table 3). [h]Each flag indicates a that a secondary criterion for significance may not be met; G, $P > 5 \times 10^{-8}$ (GWAS significant); C, Association signal was not robust against collider bias; S variant was not available in stage 2 studies for validation of stage 1 association.

stage 1 (Methods and Supplementary Table 3)[19]. Only the *RSPO3-KIAA0408* locus contains two independent variants 291 kilo bases apart, rs1892172 in *RSPO3* (MAF = 46.1%, $P_{conditional} = 4.37 \times 10^{-23}$ in the combined sexes and $P_{conditional} = 2.4 \times 10^{-20}$ in women) and rs139745911 in *KIAA0408* (MAF = 0.9%, $P_{conditional} = 3.68 \times 10^{-11}$ in combined sexes and $P_{conditional} = 1.46 \times 10^{-11}$ in women, Supplementary Table 3 and Fig. 3a). For the 33 variants within 1 Mb of previously identified WHRadjBMI SNPs, sex-combined conditional analyses identified one coding variant representing a novel independent signal in a known locus (*RREB1*; stage 1 meta-analysis, rs1334576, MAF = 44%, $P_{conditional} = 3.06 \times 10^{-7}$, Supplementary Table 4 and Fig. 3b; UKBB analysis, $P_{conditional} = 1.24 \times 10^{-8}$, Supplementary Data 7).

In summary, we identified 56 WHRadjBMI-associated coding variants in 41 independent association signals, 24 of which are new or independent of known GWAS-identified tag SNPs (either >+/−1 Mb or array-wide significant following conditional analyses) (Fig. 1). Thus, we identified 15 common and 9 low-frequency or rare novel and independent variants following conditional analyses.

Gene set and pathway enrichment analysis. To determine whether significant coding variants highlight novel or previously identified biological pathways, we applied two complementary methods, EC-DEPICT (ExomeChip Data-driven Expression-Prioritized Integration for Complex Traits)[20,21] and PASCAL[22] (Methods). For PASCAL all variants were used, for EC-DEPICT we examined only 361 variants with suggestive significance ($P < 5 \times 10^{-4}$)[10,23] from the all ancestries combined sexes analysis (which after clumping and filtering became 101 lead variants in 101 genes). We separately analyzed variants that exhibited significant sex-specific effects ($P_{sexhet} < 5 \times 10^{-4}$).

The sex-combined analyses identified 49 significantly enriched gene sets (false discovery rate, FDR < 0.05) that grouped into 25 meta-gene sets (Supplementary Note and Supplementary Data 8,9). We noted a cluster of meta-gene sets with direct relevance to metabolic aspects of obesity ('enhanced lipolysis', 'abnormal glucose homeostasis', 'increased circulating insulin level' and 'decreased susceptibility to diet-induced obesity'); we observed two significant adiponectin-related gene sets in these meta-gene sets. While

Table 2 | Association results for sex-stratified analyses. Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-7}$) in the sex-specific meta-analyses and reach Bonferroni corrected *P* value for sex heterogeneity ($P_{sexhet} < 7.14 \times 10^{-4}$)

Locus (+/− 1 Mb of a given variant)	Chr:position (GRCh37)[c]	rsID	EA	OA	Gene[d]	Amino acid change[d]	In sex-combined analyses[e]	If locus is known, nearby (< 1 Mb) published variant(s) [f]	*P* value for sex-hetero-geneity[g]	Men					Women					Other criteria for sig[i]
										N	EAF	β[h]	s.e.m.	*P*	*N*	EAF	β[h]	s.e.m.	*P*	
Variants in novel loci																				
All ancestry additive model men-only analyses																				
1	13:96665697	rs148108950	A	G	*UGGT2*	P175L	No	-	**1.5E-06**	203,009	0.006	0.130	0.024	**6.1E-08**	221,390	0.004	−0.044	0.027	1.1E-01	G
2	14:23312594	rs1042704	A	G	*MMP14*	D273N	No	-	**2.6E-04**	226,646	0.202	0.021	0.004	**2.6E-08**	250,018	0.197	0.002	0.004	6.1E-01	
All ancestry additive model women-only analyses																				
3	1:205130413	rs3851294	G	A	*DSTYK*	C641R	No	-	**9.8E-08**	225,803	0.914	−0.005	0.005	3.4E-01	249,471	0.912	0.034	0.005	**4.5E-11**	
4	2:158412701	rs55920843	T	G	*ACVR1C*	N150H	Yes	-	**1.7E-07**	210,071	0.989	0.006	0.015	7.2E-01	245,808	0.989	0.113	0.014	**1.7E-15**	
5	19:8429323	rs116843064	G	A	*ANGPTL4*	E40K	No	-	**1.3E-07**	203,098	0.981	−0.017	0.011	1.4E-01	243,351	0.981	0.064	0.011	**1.2E-09**	
Variants in previously identified loci																				
All ancestry additive model women-only analyses																				
1	1:154987704	rs141845046	C	T	*ZBTB7B*	P190S	Yes	rs905938	**7.9E-07**	226,709	0.975	0.004	0.010	6.9E-01	250,084	0.977	0.070	0.010	**2.3E-13**	
2	2:165551201	rs7607980	T	C	*COBLL1*	N941D	Yes	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	**3.0E-30**	173,600	0.880	−0.018	0.005	5.8E-04	216,636	0.878	0.062	0.005	**6.7E-39**	
3	3:129137188	rs62266958	C	T	*EFCAB12*	R197H	Yes	rs10804591	**9.3E-05**	226,690	0.937	0.018	0.006	3.1E-03	250,045	0.936	0.051	0.006	**8.1E-18**	
	3:129284818	rs2625973	A	C	*PLXND1*	L1412V	Yes		**1.6E-05**	226,650	0.736	0.005	0.003	1.9E-01	250,023	0.730	0.025	0.003	**8.2E-14**	
	3:129293256	rs2255703	T	C		M870V	Yes		**5.0E-04**	226,681	0.609	0.003	0.003	3.1E-01	250,069	0.602	0.018	0.003	**1.9E-09**	
4	4:89625427	rs1804080	G	C	*HERC3*	E946Q	Yes	rs9991328	**4.1E-06**	222,556	0.839	0.008	0.004	6.6E-02	223,877	0.837	0.034	0.004	**2.1E-16**	
	4:89668859	rs7657817	C	T	*FAM13A*	V443I	Yes		**9.6E-05**	226,680	0.816	0.006	0.004	1.5E-01	242,970	0.815	0.026	0.004	**5.9E-12**	
5	6:127476516	rs1892172	A	G	*RSPO3*	synonymous	Yes	rs11961815, rs72959041, rs1936805	**7.7E-09**	226,677	0.541	0.018	0.003	**5.6E-10**	250,034	0.545	0.042	0.003	**3.4E-48**	
	6:127767954	rs139745911[l]	A	G	*KIAA0408*	P504S	Yes		**2.0E-04**	188,079	0.010	0.057	0.017	6.8E-04	205,203	0.010	0.143	0.016	**5.9E-19**	
6	11:64031241	rs35169799	T	C	*PLCB3*	S778L	Yes	rs11231693	**1.3E-04**	226,713	0.061	0.016	0.006	9.6E-03	250,097	0.061	0.049	0.006	**6.7E-16**	
7	12:124265687	rs11057353	T	C	*DNAH10*	S228P	Yes	rs4765219, rs863750	**2.7E-08**	226,659	0.370	0.005	0.003	8.3E-02	250,054	0.376	0.029	0.003	**3.1E-22**	
	12:124330311	rs34934281	C	T		T1785M	Yes		**3.1E-08**	226,682	0.891	0.006	0.005	1.9E-01	250,066	0.887	0.043	0.005	**1.4E-20**	
	12:124427306	rs11057401	T	A	*CCDC92*	S53C	Yes		**5.5E-11**	223,324	0.701	0.013	0.003	4.3E-05	244,678	0.689	0.043	0.003	**1.0E-41**	

Abbreviations: GRCh37, human genome assembly build 37; rsID, based on dbSNP; VEP, Ensembl Variant Effect Predictor toolset; GTEx, Genotype-Tissue Expression project; *N*, sample size; EA, effect allele; OA, other allele; EAF, effect allele frequency. [a]Coding variants refer to variants located in the exons and splicing junction regions. [b]Bonferroni corrected *P* value for the number of SNPs tested for sex-heterogeneity is <7.14E-04; that is, 0.05/70 variants. [c]Variant positions are reported according to human genome assembly build 37 and their alleles are coded based on the positive strand. [d]The gene the variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on the VEP toolset and transcript abundance from GTEx database). [e]Variant was also identified as array-wide significant in the sex-combined analyses. [f]Previously published variants within +/−1 Mb are from Shungin et al.[10]. [g]*P* value for sex heterogeneity, testing for difference between women-specific and men-specific beta estimates and standard errors, was calculated using EasyStrata: Winkler et al.[11]. [h]Effect size is based on the s.d. per effect allele [i]rs139745911 in *KIAA0408* is a secondary signal in a known locus (see Supplementary Table 3). [j]Each flag indicates a that a secondary criterion for significance may not be met; G, $P > 5 \times 10^{-8}$ (GWAS significant); C, Association signal was not robust against collider bias; S, variant was not availabel in stage 2 studies for validation of stage 1 association.

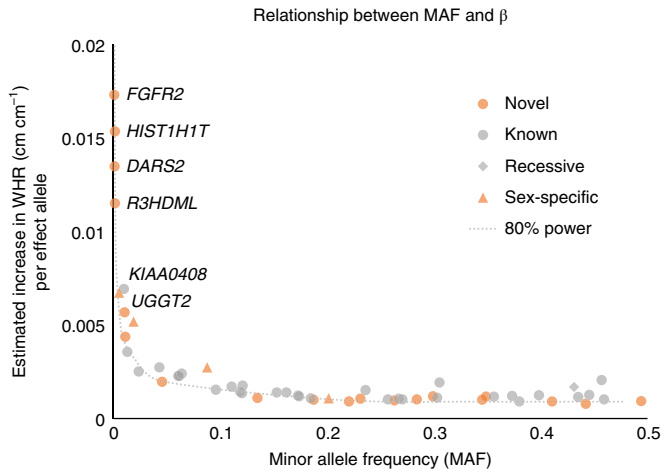


Fig. 2 | Minor allele frequency compared to estimated effect. This scatter plot displays the relationship between MAF and the estimated effect (β) for each significant coding variant in our meta-analyses. All novel WHRadjBMI variants are highlighted in orange, and variants identified only in models that assume recessive inheritance are denoted by diamonds and only in sex-specific analyses by triangles. Eighty percent power was calculated based on the total sample size in the stage 1 and 2 meta-analysis and $P = 2 \times 10^{-7}$. Estimated effects are shown in original units (cm cm^{-1}) calculated by using effect sizes in s.d. units times SD of WHR in the ARIC study (sexes combined = 0.067, men = 0.052, women = 0.080).

these pathway groups had previously been identified in the GWAS DEPICT analysis (Fig. 4), many of the individual gene sets in these meta-gene sets were not significant in the previous GWAS analysis, such as 'insulin resistance,' 'abnormal white adipose tissue physiology' and 'abnormal fat cell morphology' (Supplementary Data 8, Fig. 4 and Supplementary Fig. 16a), but represent similar biological underpinnings implied by the shared meta-gene sets. Despite their overlap with the GWAS results, these analyses highlight novel genes that fall outside known GWAS loci and with strong contributions to the substantially enriched gene sets related to adipocyte and insulin biology (for example, *MLXIPL*, *ACVR1C* and *ITIH5)* (Fig. 4).

Also, we conducted pathway analyses after excluding variants from previous WHRadjBMI analyses[10] (Supplemental Note). Seventy-five loci/genes were included in the EC-DEPICT analysis, and we identified 26 substantially enriched gene sets (13 meta-gene sets). Here, all but one gene set, 'lipid particle size', were related to skeletal biology, probably reflecting an effect on the pelvic skeleton (hip circumference), shared signaling pathways between bone and fat (such as TGF-beta) and shared developmental origin[24] (Supplementary Data 9 and Supplementary Fig. 16b). These previously identified GWAS DEPICT significant findings provide a fully independent replication of their biological relevance for WHRadjBMI.

We used PASCAL (Methods) to further distinguish between enrichment based on coding-only variant associations (this study) and regulatory-only variant associations (up to 20 kb upstream of the gene from a previous GIANT study[10]), finding 116 significantly enriched coding pathways (FDR < 0.05; Supplementary Data 10).

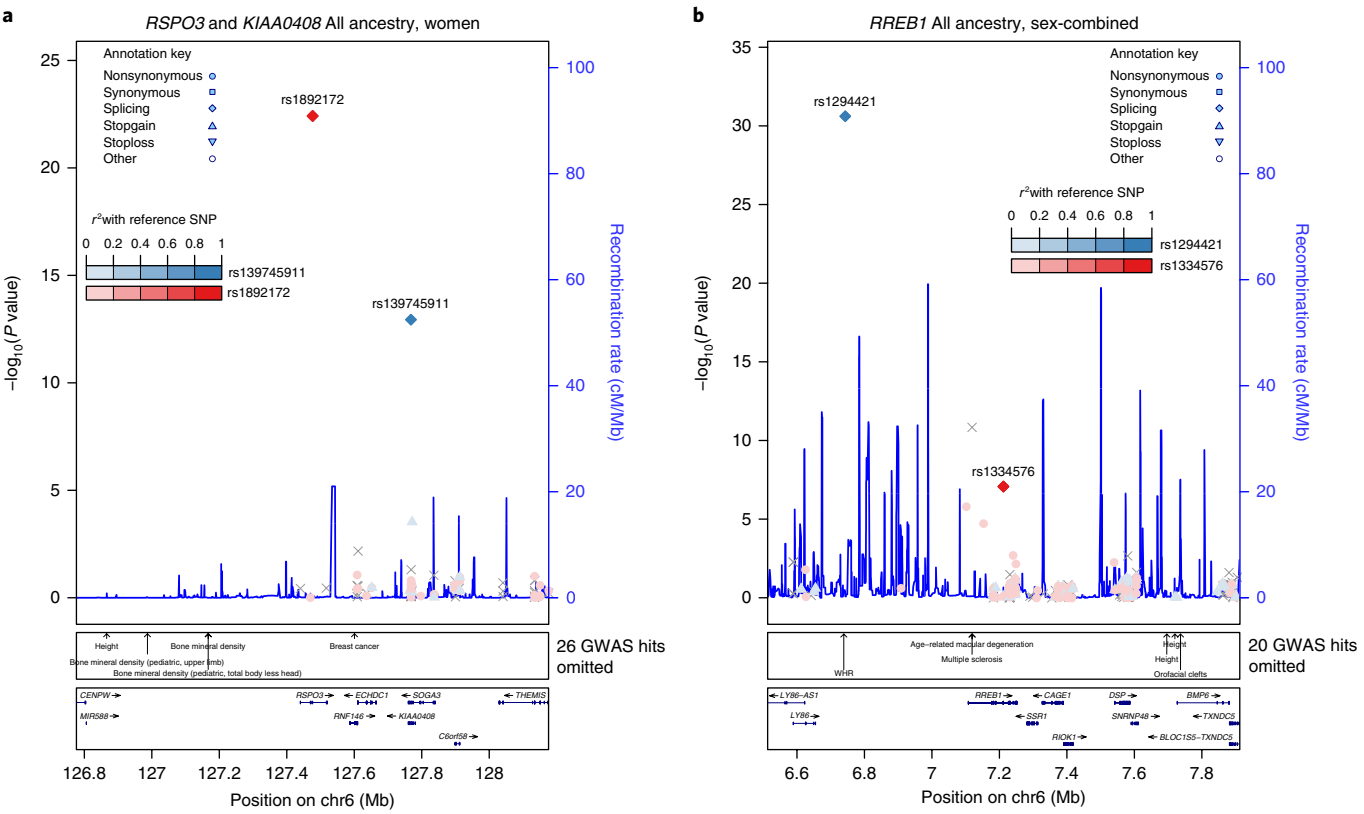


Fig. 3 | Regional association plots for known loci with novel coding signals identified by conditional analyses. Point color reflects r^2 calculated from the ARIC dataset. **a**, There are two independent variants in *RSPO3* and *KIAA0408*, based on results from the stage 1 all ancestry women ($N = 180{,}131$ for *RSPO3* and 139,056 for *KIAA0408*). **b**, We have a variant in *RREB1* that is independent of the GWAS variant rs1294410, based on results from the stage 1 all ancestry sex-combined individuals ($N = 319{,}090$; GWAS signal tagging variant rs1294421, rs1294410-rs1294421, $r^2 = 0.85$, 1000 Genomes Phase 3 version 5 all ancestries reference set).

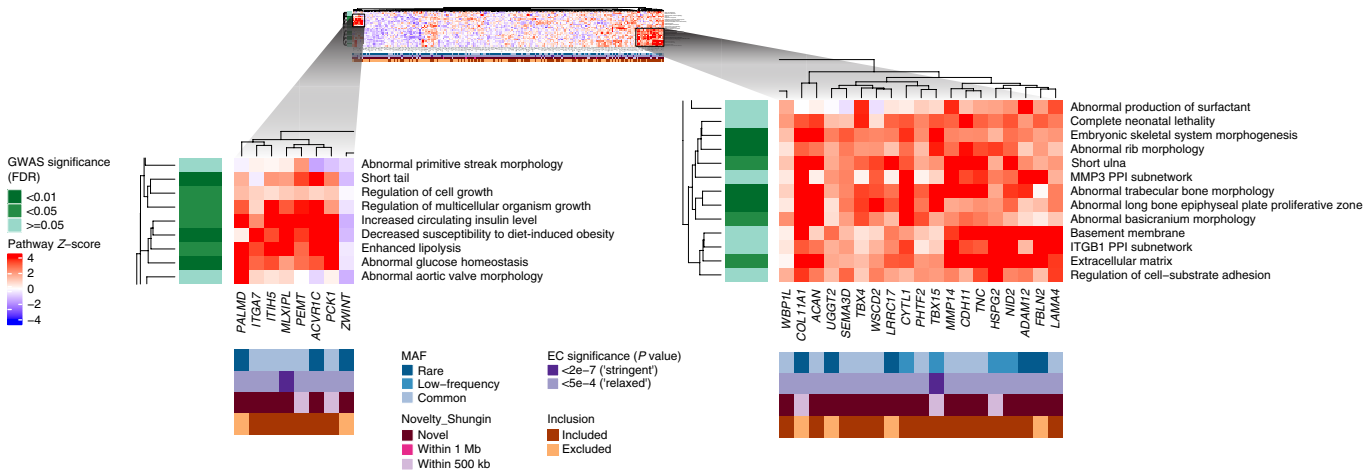


Fig. 4 | Heat maps showing DEPICT gene set enrichment results from the stage 1 all ancestry sex-combined individuals ($N = 344,369$). For any given square, the color indicates how strongly the corresponding gene (*x* axis) is predicted to belong to the reconstituted gene set (*y* axis). This value is based on the gene's *z*-score for gene set inclusion in DEPICT's reconstituted gene sets, where red indicates a higher and blue a lower *z*-score. To visually reduce redundancy and increase clarity, we chose one representative 'meta-gene set' for each group of highly correlated gene sets based on affinity propagation clustering (Methods and Supplementary Note). Heatmap intensity and DEPICT *P* values (Supplementary Data 8–9) correspond to the most substantially enriched gene set in the meta-gene set. Annotations for the genes indicate (1) the minor allele frequency of the significant ExomeChip (EC) variant (blue; if multiple variants, the lowest-frequency variant was kept), (2) whether the variant's *P* value reached array-wide significance ($<2\times10^{-7}$) or suggestive significance ($<5\times10^{-4}$) (shades of purple), (3) whether the variant was novel, overlapping 'relaxed' GWAS signals from Shungin et al.[10] (GWAS $P<5\times10^{-4}$), or overlapping 'stringent' GWAS signals (GWAS $P<5\times10^{-8}$) (pink) and (4) whether the gene was included in the gene set enrichment analysis or excluded by filters (shades of brown/orange) (Methods and Supplementary Note). Annotations for the gene sets indicate if the meta-gene set was found significant (shades of green; FDR < 0.01, < 0.05, or not significant) in the DEPICT analysis of GWAS results from Shungin et al.[10].

We also compared the coding pathways to those identified in the total previous GWAS effort (using both coding and regulatory variants) identifying a total of 158 gene sets. Forty-two gene sets were enriched in both analyses, and we found high concordance in the -log10 (*P* values) between ExomeChip and GWAS gene set enrichment (Pearson's *r* (coding versus regulatory only) = 0.38, $P<10^{-300}$; Pearson's *r* (coding versus coding + regulatory) = 0.51, $P<10^{-300}$)). Nonetheless, some gene sets were enriched *specifically* for variants in coding regions (for example, decreased susceptibility to diet-induced obesity, abnormal skeletal morphology) or unique to variants in regulatory regions (for example, transcriptional regulation of white adipocytes) (Supplementary Fig. 17).

The EC-DEPICT and PASCAL results showed a moderate but strongly significant correlation (for EC-DEPICT and the PASCAL max statistic, $r=.28$, $P=9.8\times10^{-253}$; for EC-DEPICT and the PASCAL sum statistic, $r=0.287$, $P=5.42\times10^{-272}$). Common gene sets strongly implicate a role for skeletal biology, glucose homeostasis/insulin signaling and adipocyte biology (Supplementary Fig. 18).

Cross-trait associations. To assess the clinical relevance of our identified variants with cardiometabolic, anthropometric and reproductive traits, we conducted association lookups from existing ExomeChip studies of 15 traits (Supplementary Data 11 and Supplementary Fig. 19)[21,25–29]. Variants in *STAB1* and *PLCB3* displayed the greatest number of significant associations with seven different traits ($P<9.8\times10^{-4}$, 0.05/51 variants tested). Also, these two genes cluster together with *RSPO3*, *DNAH10*, *MNS1*, *COBLL1*, *CCDC92* and *ITIH3*. The WHR-increasing alleles in this cluster exhibit a previously described pattern of increased cardiometabolic risk (for example, increased fasting insulin, two-hour glucose (TwoHGlu)) and triglycerides and decreased high-density lipoprotein cholesterol (HDL)), but also decreased BMI[30–36]. The impact of central obesity may be causal, as a 1 s.d. increase in genetic risk of central adiposity was previously associated with higher total cholesterol, triglycerides, fasting insulin and TwoHGlu and lower HDL[9].

We conducted a search in the NHGRI-EBI GWAS Catalog[37,38] to determine whether our variants are in high linkage disequilibrium ($R^2>0.7$) with variants associated with traits or diseases not covered by our cross-trait lookups (Supplementary Data 12). We identified several cardiometabolic traits (adiponectin, CHD and so on), diet/behavioral traits potentially related to obesity (carbohydrate, fat intake and so on), behavioral and neurological traits (schizophrenia, bipolar disorder and so on) and inflammatory or autoimmune diseases (Crohn's Disease, multiple sclerosis and so on).

Given the established correlation between total body-fat percentage and WHR of up to 0.483[39–41], we examined the association of our top exome variants with both total body-fat percentage and truncal fat percentage available in a sub-sample of UKBB ($N=118,160$) (Supplementary Tables 5 and 6). Seven of the common novel variants were significantly associated ($P<0.001$, 0.05/48 variants examined) with both body-fat percentage and truncal fat percentage in the sexes-combined analysis (*COBLL1*, *UHRF1BP1*, *WSCD2*, *CCDC92*, *IFI30*, *MPV17L2*, *IZUMO1*) and two with truncal fat percentage in women only (*EFCAB12*, *GDF5*). Only rs7607980 in *COBLL1* is near a known body-fat percentage GWAS locus (rs6738627; $R^2=0.1989$, distance, 6,751 base pairs, with our tag SNP)[42]. Of the nine SNPs associated with at least one of these two traits, all variants displayed much greater magnitude of effect on truncal fat percentage compared to body-fat percentage (Supplementary Fig. 20).

Previous studies have demonstrated the importance of examining common and LF/RVs in genes with mutations known to cause monogenic diseases[43,44]. Thus, we assessed enrichment of WHRadjBMI variants in monogenic lipodystrophy and/or insulin resistance genes[43,44] (Supplementary Data 13). No significant enrichment was observed, possibly due in part to the small number

Box 1 | Genes of biological interest harboring WHR-associated variants

PLXND1 (3:129284818, rs2625973, known locus). The major allele of a common non-synonymous variant in Plexin D1 (L1412V, MAF 26.7%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.0156 (0.0024), $P = 9.16 \times 10^{-11}$). *PLXND1* encodes a semaphorin class 3 and 4 receptor gene, and therefore, is involved in cell to cell signaling and regulation of growth and development for a number of different cell and tissue types, including those in the cardiovascular system, skeleton, kidneys and the central nervous system[77–81]. Mutations in this gene are associated with Moebius syndrome[82–85] and persistent truncus arteriosus[79,86]. *PLXND1* is involved in angiogenesis as part of the SEMA and VEGF signaling pathways[87–90]. *PLXND1* was implicated in the development of type 2 diabetes through its interaction with *SEMA3E* in mice. *SEMA3E* and *PLXND1* are upregulated in adipose tissue in response to diet-induced obesity, creating a cascade of adipose inflammation, insulin resistance and diabetes mellitus[81]. *PLXND1* is highly expressed in adipose (both subcutaneous and visceral) (GTeX). *PLXND1* is highly intolerant of mutations and therefore highly conserved (Supplementary Data 16). Last, our lead variant is predicted as damaging or possibly damaging for all algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

ACVR1C (2:158412701, rs55920843, novel locus). The major allele of a low-frequency non-synonymous variant in activin A receptor type 1C (rs55920843, N150H, MAF 1.1%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.0652 (0.0105), $P = 4.81 \times 10^{-10}$). *ACVR1C*, also called Activin receptor-like kinase 7 (*ALK7*), encodes a type I receptor for TGFB (Transforming Growth Factor, Beta-1), and is integral for the activation of SMAD transcription factors; therefore, *ACVR1C* plays an important role in cellular growth and differentiation[64–68], including adipocytes[68]. Mouse Acvr1c decreases secretion of insulin and is involved in lipid storage[69,72,73,91]. *ACVR1C* exhibits the highest expression in adipose tissue, but is also highly expressed in the brain (GTEx)[69–71]. Expression is associated with body fat, carbohydrate metabolism and lipids in both obese and lean individuals[70]. *ACVR1C* is moderately tolerant of mutations (ExAC constraint scores: synonymous, −0.86; non-synonymous, 1.25; LoF, 0.04 and Supplementary Data 16). Last, our lead variant is predicted as damaging for two of five algorithms examined (LRT and MutationTaster).

FGFR2 (10:123279643, rs138315382, novel locus). The minor allele of a rare synonymous variant in Fibroblast Growth Factor Receptor 2 (rs138315382, MAF 0.09%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.258 (0.049), $P = 1.38 \times 10^{-07}$). The extracellular portion of the FGFR2 protein binds with fibroblast growth factors, influencing mitogenesis and differentiation. Mutations in this gene have been associated with many rare monogenic disorders, including skeletal deformities, craniosynostosis, eye abnormalities and LADD syndrome, as well as several cancers including breast, lung and gastric cancer. Methylation of *FGFR2* is associated with high birth weight percentile[92]. *FGFR2* is tolerant of synonymous mutations, but highly intolerant of missense and loss-of-function mutations (ExAC constraint scores: synonymous, −0.9; missense, 2.74; LoF, 1.0 and Supplementary Data 16). Last, this variant is not predicted to be damaging on the basis of any of the five algorithms tested.

ANGPTL4 (19:8429323, rs116843064, novel locus). The major allele of a non-synonymous low-frequency variant in Angiopoietin Like 4 (rs116843064, E40K, EAF 98.1%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.064 (0.011) $P = 1.20 \times 10^{-09}$). *ANGPTL4* encodes a glycosylated, secreted protein containing a C-terminal fibrinogen domain. The encoded protein is induced by peroxisome proliferation activators and functions as a serum hormone that regulates glucose homeostasis, triglyceride metabolism[93,94] and insulin sensitivity[95]. Angptl4-deficient mice have hypotriglyceridemia and increased lipoprotein lipase activity, while transgenic mice overexpressing Angplt4 in the liver have higher plasma triglyceride levels and decreased ipoprotein lipase activity[96]. The major allele of rs116843064 has been previously associated with increased risk of CHD and increased triglycerides[63]. *ANGPTL4* is moderately tolerant of mutations (ExAC constraint scores: synonymous, 1.18; missense, 0.21; LoF,0.0 and Supplementary Data 16). Last, our lead variant is predicted damaging for four of five algorithms (SIFT, Polyphen 2/HDIV, Polyphen2/HVAR and MutationTaster).

RREB1 (6:7211818, rs1334576, novel association signal). The major allele of a common non-synonymous variant in the Ras responsive element binding protein 1 (rs1334576, G195R, MAF = 56%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.017 (0.002), $P = 3.9 \times 10^{-15}$). This variant is independent of the previously reported GWAS signal in the *RREB1* region (rs1294410; 6:6738752[10]). The protein encoded by this gene is a zinc finger transcription factor that binds to RAS-responsive elements (RREs) of gene promoters. It has been shown that the calcitonin gene promoter contains an RRE and that the encoded protein binds there and increases expression of calcitonin, which may be involved in Ras/Raf-mediated cell differentiation[97–99]. The ras-responsive transcription factor *RREB1* is a candidate gene for type 2 diabetes associated end-stage kidney disease[98]. This variant is highly intolerant to loss-of function (ExAC constraint score LoF, 1, Supplementary Data 16).

DAGLB (7:6449496, rs2303361, novel locus). The minor allele of a common non-synonymous variant (rs2303361, Q664R, MAF 22%) in *DAGLB* (Diacylglycerol lipase beta) is associated with increased WHRadjBMI (β (s.e.m.) = 0.0136 (0.0025), *P* value = 6.24×10^{-8}). *DAGLB* encodes a diacylglycerol (DAG) lipase that catalyzes the hydrolysis of DAG to 2-arachidonoyl-glycerol, the most abundant endocannabinoid in tissues. In the brain, DAGL activity is required for axonal growth during development and for retrograde synaptic signaling at mature synapses (2-AG)[100]. The *DAGLB* variant, rs702485 (7:6449272, $r^2 = 0.306$ and $D' = 1$ with rs2303361) has been previously associated with HDL. Pathway analyses indicate a role in the triglyceride lipase activity pathway[101]. *DAGLB* is tolerant of synonymous mutations, but intolerant of missense and loss-of-function mutations (ExAC constraint scores: synonymous, −0.76; missense, 1.07; LoF, 0.94 and Supplementary Data 16). Last, this variant is not predicted to be damaging by any of the algorithms tested.

MLXIPL (7:73012042, rs35332062 and 7:73020337, rs3812316, known locus). The major alleles of two common non-synonymous variants (A358V, MAF = 12%; Q241H, MAF = 12%) in *MLXIPL* (MLX interacting protein like) are associated with increased WHRadjBMI (β (s.e.m.) = 0.02 (0.0033), $P = 1.78 \times 10^{-9}$; β (s.e.m.) = 0.0213 (0.0034), $P = 1.98 \times 10^{-10}$). These variants are in strong linkage disequilibrium ($r^2 = 1.00$, $D' = 1.00$, 1000 Genomes CEU). This gene encodes a basic helix-loop-helix leucine zipper transcription factor of the Myc/Max/Mad superfamily. This protein forms a heterodimeric complex and binds and activates carbohydrate response element (ChoRE) motifs in the promoters of triglyceride synthesis genes in a glucose-dependent manner[74,75]. This gene is possibly involved in the growth hormone signaling pathway and lipid metabolism. The WHRadjBMI-associated

Box 1 | Genes of biological interest harboring WHR-associated variants (Continued)

variant rs3812316 in this gene has been associated with the risk of non-alcoholic fatty liver disease and coronary artery disease[74,102,103]. Furthermore, Williams–Beuren syndrome (an autosomal dominant disorder characterized by short stature, abnormal weight gain, various cardiovascular defects and mental retardation) is caused by a deletion of about 26 genes from the long arm of chromosome 7 including *MLXIPL*. *MLXIPL* is generally intolerant to variation, and therefore conserved (ExAC constraint scores: synonymous = 0.48, missense = 1.16, LoF = 0.68, Supplementary Data 16). Last, both variants reported here are predicted as possible or probably damaging by one of the algorithms tested (PolyPhen).

RAPGEF3 (12:48143315, rs145878042, novel locus). The major allele of a low-frequency non-synonymous variant in Rap Guanine-Nucleotide-Exchange Factor (GEF) 3 (rs145878042, L300P, MAF = 1.1%) is associated with increased WHRadjBMI (β (s.e.m) = 0.085 (0.010), $P = 7.15 \times 10^{-17}$). *RAPGEF3* codes for an intracellular cAMP sensor, also known as Epac (the Exchange Protein directly Activated by Cyclic AMP). Among its many known functions, RAPGEF3 regulates the ATP sensitivity of the KATP channel involved in insulin secretion[104], may be important in regulating adipocyte differentiation[105–107], plays an important role in regulating adiposity and energy balance[108]. *RAPGEF3* is tolerant of mutations (ExAC constraint scores: synonymous, −0.47; non-synonymous, 0.32; LoF, 0 and Supplementary Data 16). Last, our lead variant is predicted as damaging or possibly damaging for all five algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

TBX15 (1:119427467, rs61730011, known locus). The major allele of a low-frequency non-synonymous variant in T-box 15 (rs61730011, M460R, MAF = 4.3%) is associated with increased WHRadjBMI (β (s.e.m.) = 0.041(0.005)). T-box 15 (*TBX15*) encodes a developmental transcription factor expressed in adipose tissue, but with higher expression in visceral adipose tissue than in subcutaneous adipose tissue, and is strongly downregulated in overweight and obese individuals[109]. *TBX15* negatively controls depot-specific adipocyte differentiation and function[110] and regulates glycolytic myofiber identity and muscle metabolism[111]. *TBX15* is moderately intolerant of mutations and therefore conserved (ExAC constraint scores: synonymous, 0.42; non-synonymous, 0.65; LoF, 0.88 and Supplementary Data 16). Last, our lead variant is predicted as damaging or possibly damaging for four of five algorithms (Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

of implicated genes and the relatively small number of variants in monogenic disease-causing genes (Supplementary Fig. 21).

Genetic architecture of WHRadjBMI coding variants. We used summary statistics from our stage 1 primary meta-analysis results to estimate the phenotypic variance explained by subsets of variants across various significance thresholds ($P < 2 \times 10^{-7}$ to 0.2) and conservatively using only independent SNPs (Supplementary Table 7, Methods and Supplementary Fig. 22). For only independent coding variants that reached suggestive significance in stage 1 ($P < 2 \times 10^{-6}$), 33 SNPs explain 0.38% of the variation. The 1,786 independent SNPs with a liberal threshold of $P < 0.02$ explain 13 times more variation (5.12%), however, these large effect estimates may be subject to winner's curse. When considering all coding variants on the ExomeChip in combined sexes, 46 SNPs with a $P < 2 \times 10^{-6}$ and 5,917 SNPs with a $P < 0.02$ explain 0.51 and 13.75% of the variance in WHRadjBMI, respectively. As expected given the design of the ExomeChip, most of the variance explained is attributable to rare and low-frequency coding variants. However, for LF/RVs, those that passed significance in stage 1 explain only 0.10% of the variance in WHRadjBMI. We also estimated variance explained for the same SNPs in women and men separately and observed a greater variance explained in women compared to men ($P_{RsqDiff} < 0.002 = 0.05/21$, Bonferroni-corrected threshold) at each significance threshold considered (differences ranged from 0.24 to 0.91%).

We conducted penetrance analysis using the UKBB (both sexes combined, and men- and women only) to determine whether there is a significant accumulation of the minor allele in either the centrally obese or non-obese groups (Methods). Three rare variants (MAF ≤1%) with larger effect sizes (effect size >0.90) were included in the penetrance analysis using World Health Organization cut-offs for central obesity. Of these, one SNP (rs55920843-*ACVR1C*; $P_{sex\text{-}combined} = 9.25 \times 10^{-5}$; $P_{women} = 4.85 \times 10^{-5}$) showed a statistically significant difference in the number of carriers and non-carriers of the minor allele in the combined and female-only analysis (sex-combined obese carriers, 2.2%; non-obese carriers, 2.6%; women obese carriers, 2.1%; non-obese women carriers, 2.6%; Supplementary Table 8 and Supplementary Fig. 23).

***Drosophila* knockdown.** Considering the genetic evidence of adipose and insulin biology in determining body-fat distribution[10], and the lipid signature of the variants described herein, we examined whole-body triglyceride levels in adult *Drosophila*, a model organism in which the fat body is an organ functionally analogous to mammalian liver and adipose tissue as triglycerides are the major source of fat storage[45]. Of the 51 genes harboring our 56 significantly associated variants, we identified 27 *Drosophila* orthologs for functional follow-up analyses. We selected genes with large changes in triglyceride levels (>20% increase or >40% decrease, as chance alone is not a probable cause for changes of this magnitude) from an existing large-scale screen with ≤2 replicates per knockdown strain[45]. Two orthologs, for *PLXND1* and *DNAH10*, met these criteria and were subjected to additional knockdown experiments with ≥5 replicates using tissue-specific drivers (fat body (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdowns) (Supplementary Table 9). A significant ($P < 0.025$, 0.05/2 orthologs) increase in the total body triglyceride levels was observed in *DNAH10* ortholog knockdown strains for both the fat body and neuronal drivers. Only the neuronal driver knockdown for *PLXND1* produced a significant change in triglyceride storage. *DNAH10* and *PLXND1* both lie in previous GWAS-identified regions. Adjacent genes have been highlighted as probable candidates for the *DNAH10* association region, including *CCDC92* and *ZNF664* based on expression quantitative trait locus (eQTL) evidence. Of note, rs11057353 in *DNAH10* showed suggestive significance after conditioning on the known GWAS variants in nearby *CCDC92* (sex-combined $P_{conditional} = 7.56 \times 10^{-7}$; women-only rs11057353 $P_{conditional} = 5.86 \times 10^{-7}$, Supplementary Table 4) thus providing some evidence of multiple causal variants/genes underlying this signal. Further analyses are needed to determine whether the implicated coding variants from the current analysis are the putatively functional variants.

eQTL lookups. We examined the cis-association of variants with expression level of nearby genes in subcutaneous and visceral omental adipose, skeletal muscle and pancreas tissue from the Genotype-Tissue Expression (GTEx)[46] project, and assessed whether exome and eQTL associations implicated the same signal (Methods and Supplementary Data 14–15). The lead exome variant

was associated with expression level of the gene itself for *DAGLB, MLXIPL, CCDC92, MAPKBP1, LRRC36* and *UQCC1*. However, for *MLXIPL, MAPKBP1* and *LRRC36*, the lead variant is also associated with expression of additional nearby genes. At three additional loci, the lead exome variant is only associated with expression level of nearby genes (*HEMK1* at *C3orf18; NT5DC2, SMIM4* and *TMEM110* at *STAB1/ITIH3* and *C6orf106* at *UHRF1BP1*). Thus, although detected with a missense variant, these results are also consistent with a regulatory mechanism of effect, and the association signal may well be due to linkage disequilibrium with nearby regulatory variants.

Some of the coding genes implicated by eQTL analyses are known to be involved in adipocyte differentiation or insulin sensitivity: for example, for *MLXIPL*, the encoded carbohydrate responsive element binding protein is a transcription factor, regulating glucose-mediated induction of de novo lipogenesis in adipose tissue and expression of its *beta*-isoform in adipose tissue is positively correlated with adipose insulin sensitivity[47,48]. For *CCDC92*, the reduced adipocyte lipid accumulation on knockdown confirmed the involvement of its encoded protein in adipose differentiation[49].

Biological curation. To investigate the possible functional role of the identified variants, we conducted thorough searches of the literature and publicly available bioinformatics databases (Supplementary Data 16–17, Box 1 and Methods). Many of our novel LF/RVs are in genes that are intolerant of non-synonymous mutations (for example, *ACVR1C*, *DARS2*, *FGFR2*; ExAC Constraint Scores >0.5). Other coding variants lie in genes that are involved in glucose homeostasis (for example, *ACVR1C*, *UGGT2*, *ANGPTL4*), angiogenesis (*RASIP1*), adipogenesis (*RAPGEF3*) and lipid biology (*ANGPTL4*, *DAGLB*).

Discussion

Our analysis of coding variants from ExomeChip data in up to 476,546 individuals identified a total of 56 array-wide significant WHRadjBMI-associated variants in 41 independent association signals, including 24 newly identified (23 novel and one independent of known GWAS signals). Nine of these variants were low-frequency or rare, indicating an important role for such variants in the polygenic architecture of fat distribution. While, due to their rarity, these coding variants explain a small proportion of the trait variance at a population level, they may be more functionally tractable than non-coding variants and have a critical impact at the individual level. For instance, the association between a LF/RV (rs11209026; R381Q; MAF <5% in ExAC) located in the *IL23R* gene and multiple inflammatory diseases[50–53] led to development of new therapies targeting *IL23* and *IL12* in the same pathway[54–56]. Thus, we are encouraged that our LF/RVs displayed large effect sizes; all but one of the nine novel LF/RVs display larger effects than the 49 SNPs reported in Shungin et al. [10], and some of these effects were up to seven-fold larger than those previously reported for GWAS. This finding mirrors results for other cardiometabolic traits[57], and suggests variants of possible clinical significance with even larger effect and rarer variants will probably be detected with greater sample sizes.

We continue to observe sexual dimorphism in the genetic architecture of WHRadjBMI[11]. We identified 19 coding variants with significant sex differences, of which 16 (84%) display larger effects in women compared to men. Of the variants outside GWAS loci, we reported three (two LF/RVs) that show a significantly stronger effect in women and two (one LF/RV) that show a stronger effect in men. Genetic variants continue to explain a higher proportion of the phenotypic variation in body-fat distribution in women compared to men[10,11]. Of the novel female (*DSTYK* and *ANGPTL4*) and male (*UGGT2* and *MMP14*) specific signals, only *ANGPTL4* implicated fat distribution related biology associated with both lipid biology and cardiovascular traits (Box 1). Sexual dimorphism in fat distribution is apparent[58–60] and at sexually dimorphic loci, hormones with different levels in men and women may interact with genomic and epigenomic factors to regulate gene activity, although this remains to be tested. Dissecting the underlying molecular mechanisms of the sexual dimorphism in body-fat distribution and how it is correlated with—and causing—important comorbidities such as cardiometabolic diseases will be crucial for improved understanding of disease pathogenesis.

Overall, we observe fewer significant associations, pathways and cross-trait associations between WHRadjBMI and coding variants on the ExomeChip than Turcot et al. for BMI[25]. One reason for this may be smaller sample size ($N_{WHRadjBMI}$ = 476,546, N_{BMI} = 718,639), and thus, lower statistical power. Power is probably not the only contributing factor, as trait architecture, heritability (possibly overestimated in some phenotypes) and phenotype precision all probably contribute to our study's capacity to identify LF/RVs with large effects. Further, it is possible that the comparative lack of significant findings for WHRadjBMI may be a result of higher selective pressure against genetic predisposition to cardiometabolic phenotypes, thus rarer risk variants[61]. The ExomeChip is limited by the variants present on the chip, which was largely dictated by sequencing studies in European-ancestry populations and MAF detection criteria of ~0.012%. It is probable that through increased sample size, use of chips designed to detect variation across a range of continental ancestries, and high quality, deep imputation with large reference samples, future studies will detect additional variation from the entire allele frequency range that contributes to fat distribution.

The collected genetic and epidemiologic evidence has demonstrated that increased central adiposity is correlated with risk of type 2 diabetes and CVD, and that this association is probably causal with potential mediation through blood pressure, triglyceride-rich lipoproteins, glucose and insulin[9]. This observation yields an immediate follow-up question: which mechanisms regulate depot-specific fat accumulation and are risks for disease driven by increased visceral and/or decreased subcutaneous adipose tissue mass? Pathway analysis identified several novel pathways and gene sets related to metabolism and adipose regulation, bone growth and development and adiponectin, a hormone that has been linked to 'healthy' expansion of adipose tissue and insulin sensitivity[62]. Similarly, expression/eQTL results support the relevance of adipogenesis, adipocyte biology, and insulin signaling, supporting our previous findings for WHRadjBMI[10]. We also provide evidence suggesting known biological functions and pathways contributing to body-fat distribution (for example, diet-induced obesity, angiogenesis, bone growth/morphology and lipolysis).

The ultimate aim of genetic investigations of obesity-related traits is to identify dysregulated genomic pathways leading to obesity pathogenesis that may result in a myriad of downstream illnesses. Thus, our findings may enhance the understanding of central obesity and identify new molecular targets to avert its negative health consequences. Significant cross-trait associations are consistent with expected direction of effect for several traits; that is, the WHR-increasing allele is associated with higher values of triglycerides, DBP, fasting insulin, total cholesterol, LDL and type 2 diabetes across many significant variants. However, it is worth noting that there are some exceptions. For example, rs9469913-A in *UHRF1BP1* is associated with both increased WHRadjBMI and increased HDL. Also, we identified two variants in *MLXIPL* (rs3812316 and rs35332062), a well-known lipids-associated locus, in which the WHRadjBMI-increasing allele also increases all lipid levels, risk for hypertriglyceridemia, SBP and DBP. However, our findings show a significant and negative association with HbA1C, and nominally significant and negative associations with two-hour glucose, fasting glucose, and Type 2 diabetes, and potential negative associations with biomarkers for liver disease (for example, gamma glutamyl transpeptidase). Other notable exceptions include *ITIH3*

(negatively associated with BMI, HbA1C, LDL and SBP), *DAGLB* (positively associated with HDL), and *STAB1* (negatively associated with total cholesterol, LDL and SBP). Therefore, caution in selecting pathways for therapeutic targets is warranted; we must look beyond the effects on central adiposity to the potential cascading effects of related diseases.

A major finding from this study is the importance of lipid metabolism for body-fat distribution. In fact, pathway analyses that highlight enhanced lipolysis, cross-trait associations with circulating lipid levels, existing biological evidence from the literature, and knockdown experiments in *Drosophila*, point to novel candidate genes (*ANGPTL4*, *ACVR1C*, *DAGLB*, *MGA*, *RASIP1* and *IZUMO1*) and new candidates in known regions (*DNAH10*[10] and *MLXIPL*[14]) related to lipid biology and their role in fat storage. *ACVR1C*, *MLXIPL* and *ANGPTL4*, all of which are involved in lipid homeostasis, all are excellent candidate genes for central adiposity. Carriers of inactivating mutations in *ANGPTL4* (*Angiopoietin Like 4*), for example, display low triglycerides and low risk of coronary artery disease[63]. *ACVR1C* encodes the activin receptor-like kinase 7 protein (ALK7), a receptor for the transcription factor TGFB-1, well-known for its central role in general growth and development[64–68] and adipocyte development in particular[68]. *ACVR1C* exhibits the highest expression in adipose tissue, but is also highly expressed in the brain[69–71]. In mice, decreased activity of *ACVR1C* upregulates PPARγ and C/EBPα pathways and increases lipolysis in adipocytes, thus decreasing weight and diabetes[69,72,73]. Such activity suggests a role for ALK7 in adipose tissue signaling and a possible therapeutic target. *MLXIPL*, also important for lipid metabolism and postnatal cellular growth, encodes a transcription factor that activates triglyceride synthesis genes in a glucose-dependent manner[74,75]. The lead exome variant in *MLXIPL* is highly conserved, most probably damaging and associated with reduced *MLXIPL* expression in adipose tissue. Furthermore, in a recent longitudinal, in vitro transcriptome analysis of adipogenesis in human adipose-derived stromal cells, gene expression of *MLXIPL* was upregulated during the maturation of adipocytes, suggesting a critical role in the regulation of adipocyte size and accumulation[76]. However, given our cross-trait associations with variants in *MLXIPL* and diabetes-related traits, development of therapeutic targets must be approached cautiously.

Our 24 novel variants for WHRadjBMI highlight the importance of lipid metabolism in the genetic underpinnings of body-fat distribution. We continue to demonstrate the critical role of adipocyte biology and insulin resistance for central obesity and offer support for potentially causal genes underlying previously identified fat distribution loci. Notably, our findings offer potential new therapeutic targets for intervention in the risks associated with abdominal fat accumulation and represents a major advance in our understanding of the underlying biology and genetic architecture of central adiposity.

URLs. Ensembl ortholog database, http://www.ensembl.org; RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER; RVTEST, http://zhanxw.github.io/rvtests/; EC-DEPICT, code available at: https://github.com/RebeccaFine/obesity-ec-depict; gplots, https://CRAN.R-project.org/package=gplots.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0334-2.

Received: 17 November 2017; Accepted: 17 December 2018;
Published online: 18 February 2019

References

1. Pischon, T. et al. General and abdominal adiposity and risk of death in Europe. *N. Engl. J. Med.* **359**, 2105–2120 (2008).
2. Wang, Y., Rimm, E. B., Stampfer, M. J., Willett, W. C. & Hu, F. B. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am. J. Clin. Nutr.* **81**, 555–563 (2005).
3. Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr. Opin. Cardiol.* **23**, 591–598 (2008).
4. Snijder, M. B. et al. Associations of hip and thigh circumferences independent of waist circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am. J. Clin. Nutr.* **77**, 1192–1197 (2003).
5. Yusuf, S. et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* **366**, 1640–1649 (2005).
6. Mason, C., Craig, C. L. & Katzmarzyk, P. T. Influence of central and extremity circumferences on all-cause mortality in men and women. *Obesity.* **16**, 2690–2695 (2008).
7. Karpe, F. & Pinnick, K. E. Biology of upper-body and lower-body adipose tissue—llink to whole-body phenotypes. *Nat. Rev. Endocrinol.* **11**, 90–100 (2015).
8. Manolopoulos, K. N., Karpe, F. & Frayn, K. N. Gluteofemoral body fat as a determinant of metabolic health. *Int. J. Obes.* **34**, 949–959 (2010).
9. Emdin, C. A. et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* **317**, 626–634 (2017).
10. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
11. Winkler, T. W. et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
12. Wen, W. et al. Genome-wide association studies in east asians identify new loci for waist-hip ratio and waist circumference. *Sci. Rep.* **6**, 17958 (2016).
13. Gao, C. et al. A comprehensive analysis of common and rare variants to identify adiposity loci in hispanic Americans: the iras family study (IRASFS). *PLoS ONE* **10**, e0134649 (2015).
14. Graff, M. et al. Genome-wide physical activity interactions in adiposity—meta-analysis of 200,452 adults. *PLoS Genet.* **13**, e1006528 (2017).
15. Justice, A. E. et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* **8**, 14977 (2017).
16. Ng, M. C. Y. et al. Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet.* **13**, e1006719 (2017).
17. Aschard, H., Vilhjalmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, 329–339 (2015).
18. Day, F. R., Loh, P. R., Scott, R. A., Ong, K. K. & Perry, J. R. A robust example of collider bias in a genetic association study. *Am. J. Hum. Genet.* **98**, 392–393 (2016).
19. Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828–2829 (2014).
20. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
21. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
22. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
23. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
24. Kawai, M., de Paula, F. J. & Rosen, C. J. New insights into osteoporosis: the bone-fat connection. *J. Intern. Med.* **272**, 317–329 (2012).
25. Turcot, V. et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–41 (2018).
26. Liu, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758–1766 (2017).
27. Kraja, A. T. et al. New blood pressure-associated loci identified in meta-analyses of 475 000 individuals. *Circ. Cardiovasc. Genet.* **10**, e001778 (2017).
28. Mahajan, A. et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet.* **11**, e1004876 (2015).
29. Manning, A. et al. A low-frequency inactivating akt2 variant enriched in the finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes* **66**, 2019–2032 (2017).

30. Zhao, W. et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat. Genet.* **49**, 1450–1457 (2017).
31. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
32. Ng, M. C. et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* **10**, e1004517 (2014).
33. Mahajan, A. et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
34. Saxena, R. et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **62**, 1746–1755 (2013).
35. Cook, J. P. & Morris, A. P. Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. *Eur. J. Hum. Genet.* **24**, 1175–1180 (2016).
36. Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
37. Burdett, T. et al. The NHGRI-EBI Catalog of published genome-wide association studies. V.1.0 edn Vol. 2015 (NHGRI-EBI, 2015).
38. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
39. Lutoslawska, G. et al. Relationship between the percentage of body fat and surrogate indices of fatness in male and female Polish active and sedentary students. *J. Physiol. Anthropol.* **33**, 10 (2014).
40. Verma, M., Rajput, M., Sahoo, S. S., Kaur, N. & Rohilla, R. Correlation between the percentage of body fat and surrogate indices of obesity among adult population in rural block of Haryana. *J. Family Med. Prim. Care* **5**, 154–159 (2016).
41. Pereira, P. F. et al. Measurements of location of body fat distribution: an assessment of colinearity with body mass, adiposity and stature in female adolescents. *Rev. Paul. Pediatr.* **33**, 63–71 (2015).
42. Lu, Y. et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7**, 10495 (2016).
43. Chambers, J. C. et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat. Genet.* **40**, 716–718 (2008).
44. Nead, K. T. et al. Contribution of common non-synonymous variants in PCSK1 to body mass index variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331 175 individuals. *Hum. Mol. Genet.* **24**, 3582–3594 (2015).
45. Pospisilik, J. A. et al. *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
46. Consortium, G. T. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
47. Baraille, F., Planchais, J., Dentin, R., Guilmeau, S. & Postic, C. Integration of chrebp-mediated glucose sensing into whole body metabolism. *Physiology* **30**, 428–437 (2015).
48. Kursawe, R. et al. Decreased transcription of ChREBP-alpha/beta isoforms in abdominal subcutaneous adipose tissue of obese adolescents with prediabetes or early type 2 diabetes: associations with insulin resistance and hyperglycemia. *Diabetes* **62**, 837–844 (2013).
49. Lotta, L. A. et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
50. Cargill, M. et al. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am. J. Hum. Genet.* **80**, 273–290 (2007).
51. Hazlett, J. et al. IL-23R rs11209026 polymorphism modulates IL-17A expression in patients with rheumatoid arthritis. *Genes Immun.* 13, 282–287 (2012).
52. Karaderi, T. et al. Association between the interleukin 23 receptor and ankylosing spondylitis is confirmed by a new UK case-control study and meta-analysis of published series. *Rheumatology* **48**, 386–389 (2009).
53. Duerr, R. H. et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
54. Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A. A., Samadi, M. & Rafatpanah, H. Protective role of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: comprehensive review. *J. Immunotoxicol.* **13**, 286–300 (2016).
55. Abraham, C., Dulai, P. S., Vermeire, S. & Sandborn, W. J. Lessons learned from trials targeting cytokine pathways in patients with inflammatory bowel diseases. *Gastroenterology* **152**, 374–388 e4 (2017).
56. Molinelli, E., Campanati, A., Ganzetti, G. & Offidani, A. Biologic therapy in immune mediated inflammatory disease: basic science and clinical concepts. *Curr. Drug Saf.* **11**, 35–43 (2016).
57. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
58. Wells, J. C. Sexual dimorphism of body composition. *Best. Pract. Res. Clin. Endocrinol. Metab.* **21**, 415–430 (2007).
59. Loomba-Albrecht, L. A. & Styne, D. M. Effect of puberty on body composition. *Curr. Opin. Endocrinol. Diabetes. Obes.* **16**, 10–15 (2009).
60. Rogol, A. D., Roemmich, J. N. & Clark, P. A. Growth at puberty. *J. Adolesc. Health* **31**, 192–200 (2002).
61. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
62. Stern, J. H., Rutkowski, J. M. & Scherer, P. E. Adiponectin, leptin, and fatty acids in the maintenance of metabolic homeostasis through adipose tissue crosstalk. *Cell. Metab.* **23**, 770–784 (2016).
63. Dewey, F. E. et al. Inactivating variants in angptl4 and risk of coronary artery disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016).
64. Bondestam, J. et al. cDNA cloning, expression studies and chromosome mapping of human type I serine/threonine kinase receptor ALK7 (ACVR1C). *Cytogenet. Cell Genet.* **95**, 157–162 (2001).
65. Jornvall, H., Blokzijl, A., ten Dijke, P. & Ibanez, C. F. The orphan receptor serine/threonine kinase ALK7 signals arrest of proliferation and morphological differentiation in a neuronal cell line. *J. Biol. Chem.* **276**, 5140–5146 (2001).
66. Kim, B. C. et al. Activin receptor-like kinase-7 induces apoptosis through activation of MAPKs in a Smad3-dependent mechanism in hepatoma cells. *J. Biol. Chem.* **279**, 28458–28465 (2004).
67. Watanabe, R. et al. The MH1 domains of smad2 and smad3 are involved in the regulation of the ALK7 signals. *Biochem. Biophys. Res. Commun.* **254**, 707–712 (1999).
68. Kogame, M. et al. ALK7 is a novel marker for adipocyte differentiation. *J. Med. Invest.* **53**, 238–245 (2006).
69. Murakami, M. et al. Expression of activin receptor-like kinase 7 in adipose tissues. *Biochem. Genet.* **51**, 202–210 (2013).
70. Carlsson, L. M. et al. ALK7 expression is specific for adipose tissue, reduced in obesity and correlates to factors implicated in metabolic disease. *Biochem. Biophys. Res. Commun.* **382**, 309–314 (2009).
71. Carithers, L. J. & Moore, H. M. The genotype-tissue expression (GTEx) project. *Biopreserv. Biobank.* **13**, 307–308 (2015).
72. Yogosawa, S., Mizutani, S., Ogawa, Y. & Izumi, T. Activin receptor-like kinase 7 suppresses lipolysis to accumulate fat in obesity through downregulation of peroxisome proliferator-activated receptor gamma and C/EBPalpha. *Diabetes* **62**, 115–123 (2013).
73. Yogosawa, S. & Izumi, T. Roles of activin receptor-like kinase 7 signaling and its target, peroxisome proliferator-activated receptor gamma, in lean and obese adipocytes. *Adipocyte* **2**, 246–250 (2013).
74. Seifi, M., Ghasemi, A., Namipashaki, A. & Samadikuchaksaraei, A. Is C771G polymorphism of MLX interacting protein-like (MLXIPL) gene a novel genetic risk factor for non-alcoholic fatty liver disease? *Cell Mol. Biol.* **60**, 37–42 (2014).
75. Cairo, S., Merla, G., Urbinati, F., Ballabio, A. & Reymond, A. WBSCR14, a gene mapping to the Williams–Beuren syndrome deleted region, is a new member of the Mlx transcription factor network. *Hum. Mol. Genet.* **10**, 617–627 (2001).
76. Ambele, M. A., Dessels, C., Durandt, C. & Pepper, M. S. Genome-wide analysis of gene expression during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene expression during adipocyte differentiation. *Stem Cell Res.* **16**, 725–734 (2016).
77. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
78. Toyofuku, T. et al. Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1. *EMBO J.* **26**, 1373–1384 (2007).
79. Gitler, A. D., Lu, M. M. & Epstein, J. A. PlexinD1 and semaphorin signaling are required in endothelial cells for cardiovascular development. *Dev. Cell.* **7**, 107–116 (2004).
80. Luchino, J. et al. Semaphorin 3E suppresses tumor cell death triggered by the plexin D1 dependence receptor in metastatic breast cancers. *Cancer Cell.* **24**, 673–685 (2013).
81. Shimizu, I. et al. Semaphorin3E-induced inflammation contributes to insulin resistance in dietary obesity. *Cell. Metab.* **18**, 491–504 (2013).
82. Verzijl, H. T., van der Zwaag, B., Cruysberg, J. R. & Padberg, G. W. Mobius syndrome redefined: a syndrome of rhombencephalic maldevelopment. *Neurology* **61**, 327–333 (2003).
83. Verzijl, H. T., van der Zwaag, B., Lammens, M., ten Donkelaar, H. J. & Padberg, G. W. The neuropathology of hereditary congenital facial palsy vs Mobius syndrome. *Neurology* **64**, 649–653 (2005).

84. Fujita, M., Reinhart, F. & Neutra, M. Convergence of apical and basolateral endocytic pathways at apical late endosomes in absorptive cells of suckling rat ileum in vivo. *J. Cell Sci.* **97**(Pt 2), 385–394 (1990).
85. Briegel, W. Neuropsychiatric findings of mobius sequence—a review. *Clin. Genet.* **70**, 91–97 (2006).
86. Ta-Shma, A. et al. Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene. *Am. J. Med. Genet. A.* **161A**, 3115–3120 (2013).
87. Mazzotta, C. et al. Plexin-D1/Semaphorin 3E pathway may contribute to dysregulation of vascular tone control and defective angiogenesis in systemic sclerosis. *Arthritis Res. Ther.* **17**, 221 (2015).
88. Yang, W. J. et al. Semaphorin-3C signals through Neuropilin-1 and PlexinD1 receptors to inhibit pathological angiogenesis. *EMBO Mol. Med.* **7**, 1267–1284 (2015).
89. Zygmunt, T. et al. Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy receptor sFlt1. *Dev. Cell.* **21**, 301–314 (2011).
90. Kim, J., Oh, W. J., Gaiano, N., Yoshida, Y. & Gu, C. Semaphorin 3E-Plexin-D1 signaling regulates VEGF function in developmental angiogenesis via a feedback mechanism. *Genes Dev.* **25**, 1399–1411 (2011).
91. Bertolino, P. et al. Activin B receptor ALK7 is a negative regulator of pancreatic beta-cell function. *Proc. Natl Acad. Sci. USA* **105**, 7246–7251 (2008).
92. Haworth, K. E. et al. Methylation of the FGFR2 gene is associated with high birth weight centile in humans. *Epigenomics* **6**, 477–491 (2014).
93. Chi, X. et al. Angiopoietin-like 4 modifies the interactions between lipoprotein lipase and its endothelial cell transporter GPIHBP1. *J. Biol. Chem.* **290**, 11865–11877 (2015).
94. Catoire, M. et al. Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise. *Proc. Natl Acad. Sci. USA* **111**, E1043–E1052 (2014).
95. van Raalte, D. H. et al. Angiopoietin-like protein 4 is differentially regulated by glucocorticoids and insulin in vitro and in vivo in healthy humans. *Exp. Clin. Endocrinol. Diabetes.* **120**, 598–603 (2012).
96. Koster, A. et al. Transgenic angiopoietin-like (angptl)4 overexpression and targeted disruption of angptl4 and angptl3: regulation of triglyceride metabolism. *Endocrinology* **146**, 4943–4950 (2005).
97. Thiagalingam, A. et al. RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas. *Mol. Cell. Biol.* **16**, 5335–5345 (1996).
98. Bonomo, J. A. et al. The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum. Mol. Genet.* **23**, 6441–6447 (2014).
99. Thiagalingam, A., Lengauer, C., Baylin, S. B. & Nelkin, B. D. RREB1, a ras responsive element binding protein, maps to human chromosome 6p25. *Genomics* **45**, 630–632 (1997).
100. Bisogno, T. et al. Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation of endocannabinoid signaling in the brain. *J. Cell Biol.* **163**, 463–468 (2003).
101. Global Lipids Genetics, C. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
102. Kooner, J. S. et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.* **40**, 149–151 (2008).
103. Pan, L. A. et al. G771C polymorphism in the mlxipl gene is associated with a risk of coronary artery disease in the chinese: a case-control study. *Cardiology* **114**, 174–178 (2009).
104. Kang, G., Leech, C. A., Chepurny, O. G., Coetzee, W. A. & Holz, G. G. Role of the cAMP sensor Epac as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1 cells. *J. Physiol.* **586**, 1307–1319 (2008).
105. Ji, Z., Mei, F. C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte differentiation. *Front Biosci.* **2**, 392–398 (2010).
106. Martini, C. N., Plaza, M. V. & Vila Mdel, C. PKA-dependent and independent cAMP signaling in 3T3-L1 fibroblasts differentiation. *Mol. Cell. Endocrinol.* **298**, 42–47 (2009).
107. Petersen, R. K. et al. Cyclic AMP (cAMP)-mediated stimulation of adipocyte differentiation requires the synergistic action of Epac- and cAMP-dependent protein kinase-dependent processes. *Mol. Cell. Biol.* **28**, 3804–3816 (2008).
108. Yan, J. et al. Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol. Cell. Biol.* **33**, 918–926 (2013).
109. Gesta, S. et al. Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc. Natl Acad. Sci. USA* **103**, 6676–6681 (2006).
110. Gesta, S. et al. Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. *Proc. Natl Acad. Sci. USA* **108**, 2771–2776 (2011).
111. Lee, K. Y. et al. Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism. *Nat. Commun.* **6**, 8054 (2015).

Acknowledgements

This work was primarily supported through funding from the National Institute of Health (NIH): 1K99HL130580, R01-DK089256, 2R01HD057194, U01HG007416, R01DK101855, T32 HL007055, KL2TR001109; and the American Heart Association (AHA): 13POST16500011 and 13GRNT16490017. Co-author Y. Jia recently passed away while this work was in process. This study was completed as part of the Genetic Investigation of ANtropometric Traits (GIANT) Consortium. This research has been conducted using the UK Biobank resource. A full list of acknowledgements is provided in the Supplementary Data 18.

Author contributions

Writing Group: L.A.C., R.S.F., T.M.F., M. Graff, H.M.H., J.N.H., A.E.J., T.K., Z.K., C.M.L., R.J.F.L., Y.L., K.E.N., V.T., K.L.Y.

Data preparation group: T.A., I.B.B., T.E., S.F., M. Graff, H.M.H., A.E.J., T.K., D.J.L., K.S.L., A.E.L., R.J.F.L., Y.L., E. Marouli, N.G.D.M., C.M.-G, P.M., M.C.Y.N., M.A.R., S.S., C.S., K. Stirrups, V.T., S.V., S.M.W., T.W.W., K.L.Y., X.Z.

WHR meta-analyses: P.L.A., H.M.H., A.E.J., T.K., M. Graff, C.M.L., R.J.F.L., K.E.N., V.T., K.L.Y.

Pleiotropy working group: G.A., M. Boehnke, J.P.C., P.D., F.D., J.C.F., H.M.H., H.K., H.M.H., A.E.J., C.M.L., D.J.L., R.J.F.L., A. Mahajan, E. Marouli, G.M., M.I.M., P.B.M., G.M.P., J.R.B.P., K.S.R., X.S., S.W., J.W., C.J.W.

Phenome-wide association studies: L. Bastarache, J.C.D., A.G., A. Mahajan, M.I.M.

Gene-set enrichment analyses: S.B., R.S.F., J.N.H., Z.K., D.L., T.H.P., T.F.V.

eQTL analyses: C.K.R., Y.L., K.L.M.

Monogenic and syndromic gene enrichment analyses: H.M.H., A.K.M.

Fly Obesity Screen: A. Lempradl, J.A. Pospisilik.

Overseeing of contributing studies and consortia: (1958 Birth Cohort) P.D.; (Airwave) P.E.; (AMC PAS) G.K.H.; (Amish) J.R.O.; (ARIC) E.B.; (ARIC, Add Health) K.E.N.; (BRAVE) E.D.A., R.C.; (BRIGHT) P.B.M.; (CARDIA) M.F., P.J.S.; (Cebu Longitudinal Health and Nutrition Survey) K.L.M.; (CHD Exome + Consortium) A.S.B., J.M.M.H., D.F.R., J.D.; (CHES) R.V.; (Clear/eMERGE (Seattle)) G.P.J.; (CROATIA_Korcula) V.V., O. Polasek, I.R.; (deCODE) K. Stefansson, U.T.; (DHS) D.W.B.; (DIACORE) C.A.B.; (DPS) J.T., J. Lindström, M.U.; (DRSEXTRA) T.A.L., R.R.; (EFSOCH) A.T.H., T.M.F.; (EGCUT) T.E.; (eMERGE (Seattle)) E.B.L.; (EPIC-Potsdam) M.B.S., H.B.; (EpiHealth) E.I., P.W.F.; (EXTEND) A.T.H., T.M.F.; (Family Heart Study) I.B.B.; (Fenland, EPIC) R.A.S.; (Fenland, EPIC, InterAct) N.J.W., C.L.; (FINRISK) S. Männistö; (FINRISK 2007 (T2D)) P.J., V. Salomaa; (Framingham Heart Study) L.A.C.; (FUSION) M. Boehnke, F.S.C.; (FVG) P.G.; (Generation Scotland) C.H., B.H.S.; (Genetic Epidemiology Network of Arteriopathy (GENOA)) S.L.R.K.; (GRAPHIC) N.J.S.; (GSK-STABILITY) D.M.W., L.W., H.D.W.; (Health) A. Linneberg; (HELIC MANOLIS) E.Z., G.D.; (HELIC Pomak) E.Z., G.D.; (HUNT-MI) K.H., C.J.W.; (Inter99) T.H., T.J.; (IRASFS) L.E.W., E.K.S.; (Jackson Heart Study (JHS)) J.G.W.; (KORA S4) K. Strauch, I.M.H.; (Leipzig-Adults) M. Blüher, P. Kovacs.; (LOLIPOP-Exome) J.C.C., J.S.K.; (LOLIPOP-OmniEE) J.C.C., J.S.K.; (MESA) J.I.R., X.G.; (METSIM) J.K., M.L.; (MONICA-Brianza) G.C.; (Montreal Heart Institute Biobank (MHIBB)) M.P.D., G.L., S.d.D., J.C.T.; (MORGAM Central Laboratory) M.P.; (MORGAM Data Centre) K.K.; (OBB) F. Karpe; (PCOS) A.P.M., C.M.L.; (PIVUS) C.M.L., L.L.; (PRIME—Belfast) F. Kee; (PRIME—Lille) P.A.; (PRIME—Strasbourg) M.M.; (PRIME—Toulouse) J.F.; (PROMIS) D.S.; (quality control) M.A.R.; (RISC) B.B., E.F., M.W.; (Rotterdam Study I) A.G.U., M.A.I.; (SEARCH) A.M.D.; (SHIP/SHIP-Trend) M.D.; (SIBS) D.F.E.; (SOLID TIMI-52) D.M.W.; (SORBS) A.P.M., M.S., A. Tönjes; (The Mount Sinai BioMe Biobank) E.P.B., R.J.F.L.; (The NEO Study) D.O.M.K.; (The NHAPC Study, The GBTDS Study) X.L.; (The Western Australian Pregnancy Cohort (Raine) Study) C.E.P., S. Macgregor; (TwinsUK) T.D.S.; (ULSAM) A.P.M.; (Vejle Biobank) I.B., C.C., O. Pedersen; (WGHS) D.I.C., P.M.R.; (Women's Health Initiative) P.L.A.; (WTCCC-UKT2D) M.I.M., K.R.O.; (YFS) T.L., O.T.R.

Genotyping of contributing studies and consortia: (1958 Birth Cohort) K.E.S.; (Airwave) E.E., M.S.L.; (AMC PAS) S.S.; (Amish) L.M.Y.A., J.A. Perry; (ARIC) E.W.D., M.L.G.; (BBMRI-NL) S.H.V., L. Broer, C.M.v.D., P.I.W.d.B.; (BRAVE) E.D.A.; (Cambridge Cancer Studies) J.G.D.; (CARDIA) M.F.; (CHD Exome + Consortium) A.S.B., J.M.M.H., D.F.R., J.D., R.Y.; (Clear/eMERGE (Seattle)) G.P.J.; (CROATIA_Korcula) V.V.; (DIACORE) C.A.B., M. Gorski; (DPS) A.U.J., J. Lindström; (DRSEXTRA) P. Komulainen; (EGCUT) T.E.; (EPIC-Potsdam) M.B.S., K.M.; (EpiHealth) E.I., P.W.F.; (Family Heart Study) K.D.T.; (Fenland, EPIC) R.A.S.; (Fenland, EPIC, InterAct) N.J.W., C.L.; (FUSION) N.N.; (FVG) I.G., A. Morgan; (Generation Scotland) C.H.; (Genetic Epidemiology Network of Arteriopathy (GENOA)) S.L.R.K., J.A.S.; (GRAPHIC) N.J.S.; (GSK-STABILITY) D.M.W.; (Health) J.B.J.; (HELIC MANOLIS) L. Southam; (HELIC Pomak) L. Southam; (Inter99) T.H., N.G.; (KORA) M.M.N.; (KORA S4) K. Strauch, H.G.; (Leipzig-Adults) A. Mahajan; (LOLIPOP-Exome) J.C.C., J.S.K.; (LOLIPOP-OmniEE) J.C.C., J.S.K.; (MESA) J.I.R., Y.D.I.C., K.D.T.; (METSIM) J.K., M.L.; (Montreal Heart Institute Biobank (MHIBB)) M.P.D.; (OBB) F. Karpe; (PCOS) A.P.M.; (PIVUS) C.M.L.; (Rotterdam Study I) A.G.U., C.M.G., F.R.; (SDC) J.M.J., H.V.; (SEARCH) A.M.D.; (SOLID TIMI-52) D.M.W.; (SORBS) A.P.M.; (The Mount Sinai BioMe Biobank) E.P.B., R.J.F.L., Y.L., C.S.; (The NEO Study) R.L.G.; (The NHAPC study, The GBTDS study) X.L., H.L., Y.H.; (The Western Australian Pregnancy Cohort (Raine) Study) C.E.P., S. Macgregor; (TUDR) Z.A.; (TwinsUK) A.P.M.; (ULSAM) A.P.M.; (WGHS) D.I.C., A.Y.C.; (Women's Health Initiative) A.P.R.; (WTCCC-UKT2D) M.I.M.; (YFS) T.L., L.P.L.

Phenotyping of contributing studies and consortia: (Airwave) E.E.; (AMC PAS) S.S.; (Amish) L.M.Y.A.; (ARIC) E.W.D.; (ARIC, Add Health) K.E.N.; (BBMRI-NL) S.H.V.; (BRAVE) E.D.A.; (BRIGHT) M.J.C.; (CARL) A. Robino, G.G.; (Cebu Longitudinal

Health and Nutrition Survey) N.R.L.; (CHES) R.V., M.T.; (Clear/eMERGE (Seattle)) G.P.J., A.A.B.; (CROATIA_Korcula) O. Polasek, I.R.; (DIACORE) C.A.B., B.K.K.; (DPS) A.U.J., J. Lindström; (EFSOCH) A.T.H.; (EGCUT) E. Mihailov; (EPIC-Potsdam) H.B.; (EpiHealth) E.I.; (EXTEND) A.T.H.; (Family Heart Study) M.F.F.; (Fenland, EPIC, InterAct) N.J.W.; (FIN-D2D 2007) L.M., M.V.; (FINRISK) S. Männistö; (FINRISK 2007 (T2D)) P.J., H.M.S.; (Framingham Heart Study) C.S.F.; (Generation Scotland) C.H., B.H.S.; (Genetic Epidemiology Network of Arteriopathy (GENOA)) S.L.R.K., J.A.S.; (GRAPHIC) N.J.S.; (GSK-STABILITY) L.W., H.D.W.; (Health) A. Linneberg, B.H.T.; (HELIC MANOLIS) L. Southam, A.E.F., E.T.; (HELIC Pomak) L. Southam, A.E.F., M.K.; (HUNT-MI) K.H., O.L.H.; (Inter99) T.J., N.G.; (IRASFS) L.E.W., B.K.; (KORA) M.M.N.; (BBMRI-NL) K.M.A.S.; (Leipzig-Adults) M. Blüher, P. Kovacs; (LOLIPOP-Exome) J.C.C., J.S.K.; (LOLIPOP-OmniEE) J.C.C., J.S.K.; (MESA) M.A.; (Montreal Heart Institute Biobank (MHIBB)) G.L., K.S.L., V.T.; (MORGAM Data Centre) K.K.; (OBB) F. Karpe, M.N.; (PCOS) C.M.L.; (PIVUS) L.L.; (PRIME—Belfast) F. Kee; (PRIME—Lille) P.A.; (PRIME—Strasbourg) M.M.; (PRIME—Toulouse) J.F.; (RISC) B.B., E.F.; (Rotterdam Study I) M.A.I., C.M.G., F.R., M.C.Z.; (SHIP/SHIP-Trend) N.F.; (SORBS) M.S., A. Tönjes; (The Mount Sinai BioMe Biobank) E.P.B., Y.L., C.S.; (The NEO Study) R.d.M.; (The NHAPC study, The GBTDS study) X.L., H.L., L. Sun, F.W.; (The Western Australian Pregnancy Cohort (Raine) Study) C.E.P.; (TUDR) Y.J.H., W.J.L.; (TwinsUK) T.D.S., K.S.S.; (ULSAM) V.G.; (WGHS) D.I.C., P.M.R.; (Women's Health Initiative) A.P.R.; (WTCCC-UKT2D) M.I.M., K.R.O.; (YFS) T.L., O.T.R.

Data analysis of contributing studies and consortia: (1958 Birth Cohort) K.E.S., I.N.; (Airwave) E.E., M.S.L.; (AMC PAS) S.S.; (Amish) J.R.O., L.M.Y.A., J.A. Perry; (ARIC, Add Health) K.E.N., K.L.Y., M. Graff; (BBMRI-NL) L. Broer; (BRAVE) R.C., D.S.A.; (BRIGHT) H.R.W.; (Cambridge Cancer Studies) J.G.D., A.E., D.J.T.; (CARDIA) C.E.L, M.F., L.-A.L.; (CARL) A. Robino, D.V.; (Cebu Longitudinal Health and Nutrition Survey) Y.W.; (CHD Exome + Consortium) A.S.B., J.M.M.H., D.F.R., R.Y., P.S.; (CHES) Y.J.; (CROATIA_Korcula) V.V.; (deCODE) V. Steinthorsdottir, G.T.; (DHS) A.J.C., P.M., M.C.Y.N.; (DIACORE) C.A.B., M. Gorski; (EFSOCH) H.Y.; (EGCUT) T.E., R.M.; (eMERGE (Seattle)) D.S.C.; (ENDO) T.K.; (EPIC) J.H.Z.; (EPIC-Potsdam) K.M.; (EpiHealth) S.G.; (EXTEND) H.Y.; (Family Heart Study) M.F.F.; (Fenland) J.Luan.; (Fenland, EPIC) R.A.S.; (Fenland, InterAct) S.M.W.; (FIN-D2D 2007) M.V., L.M.; (Finrisk Extremes and quality control) S.V.; (Framingham Heart Study) C.T.L., N.L.H.C.; (FVG) I.G.; (Generation Scotland) C.H., J.M.; (Genetic Epidemiology Network of Arteriopathy (GENOA)) L.F.B.; (GIANT-Analyst) A.E.J.; (GRAPHIC) N.J.S., N.G.D.M., C.P.N.; (GSK-STABILITY) D.M.W., A.J.S.; (Health) J.B.J.; (HELIC MANOLIS) L. Southam; (HELIC Pomak) L. Southam; (HUNT-MI) W. Zhou; (Inter99) N.G.; (IRASFS) B.K.; (Jackson Heart Study (JHS)) L.A.L., J. Li; (KORA S4) T.W.W.; (BBMRI-NL) K.M.A.S.; (Leipzig-Adults) A. Mahajan; (LOLIPOP-Exome) J.C.C., J.S.K., W. Zhang; (LOLIPOP-OmniEE) J.C.C., J.S.K., Weihua Zhang; (MESA) J.I.R., X.G., J.Y.; (METSIM) X.S.; (Montreal Heart Institute Biobank (MHIBB)) J.C.T., G.L., K.S.L., V.T.; (OBB) A. Mahajan; (PCOS) A.P.M., T.K.; (PIVUS) N.R.R.; (PROMIS) A. Rasheed, W. Zhao; (quality control GoT2D/T2D-GENES (FUSION, METSIM and so on)) A.E.L.; (RISC) H.Y.; (Rotterdam Study I) C.M.G., F.R.; (SHIP/SHIP-Trend) A. Teumer; (SOLID TIMI-52) D.M.W., A.J.S.; (SORBS) A.P.M; (The Mount Sinai BioMe Biobank) Y.L., C.S.; (The NEO Study) R.L.G.; (The NHAPC study, The GBTDS study) X.L., H.L., Y.H., H.Z.; (The Western Australian Pregnancy Cohort (Raine) Study) C.A.W.; (UK Biobank) A.R.W.; (ULSAM) A.P.M., A. Mahajan; (WGHS) D.I.C., A.Y.C.; (Women's Health Initiative) P.L.A., J.H.; (WTCCC-UKT2D) W.G.; (YFS) L.P.L.

Competing interests

The authors declare the following competing interests: A.S.B. holds interest in AstraZeneca, Biogen, Bioverativ, Merck, Novartis and Pfizer. A.S.C. and C.S.F. are current employees of Merck. Authors affiliated with deCODE (V.St., G.T., U.T. and K.S.) are employed by deCODE Genetics/Amgen, Inc. H.D.W. has the following financial and non-financial competing interests to declare: research grants: Sanofi Aventis; Eli Lilly; NIH; Omthera Pharmaceuticals, Pfizer, Elsai Inc. AstraZeneca; DalCor and Services; lecture fees: Sanofi Aventis; advisory boards: Acetelion, Sirtex, CSL Boehring. J.D. has received grants from AstraZeneca, Biogen, Merck, Novartis and Pfizer. L.M.Y.A. and R.A.S. are employee stock holders of GlaxoSmithKline. M.P.D. received honoraria and holds minor equity in Dalcor. V.S. has participated in a conference trip sponsored by Novo Nordisk.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0334-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.E.N. or C.M.L.

Anne E. Justice[1,2,225], Tugce Karaderi[3,4,225], Heather M. Highland[1,5,225], Kristin L. Young[1,225], Mariaelisa Graff[1,225], Yingchang Lu[6,7,8,225], Valérie Turcot[9,225], Paul L. Auer[10], Rebecca S. Fine[11,12,13], Xiuqing Guo[14], Claudia Schurmann[7,8], Adelheid Lempradl[15], Eirini Marouli[16], Anubha Mahajan[3], Thomas W. Winkler[17], Adam E. Locke[18,19], Carolina Medina-Gomez[20,21], Tõnu Esko[11,13,22], Sailaja Vedantam[11,12,13], Ayush Giri[23], Ken Sin Lo[9,23], Tamuno Alfred[7], Poorva Mudgal[24], Maggie C. Y. Ng[24,25], Nancy L. Heard-Costa[26,27], Mary F. Feitosa[28], Alisa K. Manning[11,29,30], Sara M. Willems[31], Suthesh Sivapalaratnam[30,32,33], Goncalo Abecasis[18,34], Dewan S. Alam[35], Matthew Allison[36], Philippe Amouyel[37,38,39], Zorayr Arzumanyan[14], Beverley Balkau[40], Lisa Bastarache[41], Sven Bergmann[42,43], Lawrence F. Bielak[44], Matthias Blüher[45,46], Michael Boehnke[18], Heiner Boeing[47], Eric Boerwinkle[5,48], Carsten A. Böger[49], Jette Bork-Jensen[50], Erwin P. Bottinger[7], Donald W. Bowden[24,25,51], Ivan Brandslund[52,53], Linda Broer[21], Amber A. Burt[54], Adam S. Butterworth[55,56], Mark J. Caulfield[16,57], Giancarlo Cesana[58], John C. Chambers[59,60,61,62,63], Daniel I. Chasman[11,64,65,66], Yii-Der Ida Chen[14], Rajiv Chowdhury[55], Cramer Christensen[67], Audrey Y. Chu[65], Francis S. Collins[68], James P. Cook[69], Amanda J. Cox[24,25,70], David S. Crosslin[71], John Danesh[55,56,72,73], Paul I. W. de Bakker[74,75], Simon de Denus[9,76], Renée de Mutsert[77], George Dedoussis[78], Ellen W. Demerath[79], Joe G. Dennis[80], Josh C. Denny[41], Emanuele Di Angelantonio[55,56,73], Marcus Dörr[81,82], Fotios Drenos[83,84,85], Marie-Pierre Dubé[9,86], Alison M. Dunning[87], Douglas F. Easton[80,87], Paul Elliott[88], Evangelos Evangelou[61,89], Aliki-Eleni Farmaki[78], Shuang Feng[18], Ele Ferrannini[90,91], Jean Ferrieres[92], Jose C. Florez[11,29,30], Myriam Fornage[93], Caroline S. Fox[27], Paul W. Franks[94,95,96], Nele Friedrich[97], Wei Gan[3], Ilaria Gandin[98], Paolo Gasparini[99,100], Vilmantas Giedraitis[101], Giorgia Girotto[99,100], Mathias Gorski[17,49], Harald Grallert[102,103,104], Niels Grarup[50], Megan L. Grove[5], Stefan Gustafsson[105], Jeff Haessler[106], Torben Hansen[50], Andrew T. Hattersley[107], Caroline Hayward[108], Iris M. Heid[17,109], Oddgeir L. Holmen[110], G. Kees Hovingh[111], Joanna M. M. Howson[55], Yao Hu[112], Yi-Jen Hung[113,114], Kristian Hveem[110,115], M. Arfan Ikram[20,116,117], Erik Ingelsson[105,118], Anne U. Jackson[18], Gail P. Jarvik[54,119], Yucheng Jia[14], Torben Jørgensen[120,121,122], Pekka Jousilahti[123], Johanne M. Justesen[50], Bratati Kahali[124,125,126,127], Maria Karaleftheri[128], Sharon L. R. Kardia[44], Fredrik Karpe[129,130], Frank Kee[131], Hidetoshi Kitajima[3], Pirjo Komulainen[132], Jaspal S. Kooner[60,62,63,133], Peter Kovacs[45], Bernhard K. Krämer[134], Kari Kuulasmaa[123], Johanna Kuusisto[135], Markku Laakso[135], Timo A. Lakka[132,136,137], David Lamparter[42,43,138], Leslie A. Lange[139], Claudia Langenberg[31], Eric B. Larson[54,140,141], Nanette R. Lee[142,143], Wen-Jane Lee[144,145], Terho Lehtimäki[146,147], Cora E. Lewis[148], Huaixing Li[112], Jin Li[149], Ruifang Li-Gao[77], Li-An Lin[93], Xu Lin[112], Lars Lind[150], Jaana Lindström[123], Allan Linneberg[122,151,152], Ching-Ti Liu[153], Dajiang J. Liu[154], Jian'an Luan[31], Leo-Pekka Lyytikäinen[146,147], Stuart MacGregor[155], Reedik Mägi[22], Satu Männistö[123], Gaëlle Marenne[72], Jonathan Marten[108], Nicholas G. D. Masca[156,157], Mark I. McCarthy[3,129,130], Karina Meidtner[102,158], Evelin Mihailov[22], Leena Moilanen[159], Marie Moitry[160,161], Dennis O. Mook-Kanamori[77,162], Anna Morgan[99], Andrew P. Morris[3,69], Martina Müller-Nurasyid[109,163,164], Patricia B. Munroe[16,57], Narisu Narisu[68], Christopher P. Nelson[156,157], Matt Neville[129,130], Ioanna Ntalla[16], Jeffrey R. O'Connell[165], Katharine R. Owen[129,130], Oluf Pedersen[50], Gina M. Peloso[153], Craig E. Pennell[166,167], Markus Perola[123,168], James A. Perry[165], John R. B. Perry[31], Tune H. Pers[50,169], Ailith Ewing[80], Ozren Polasek[170,171], Olli T. Raitakari[172,173], Asif Rasheed[174], Chelsea K. Raulerson[175], Rainer Rauramaa[132,136], Dermot F. Reilly[176], Alex P. Reiner[106,177], Paul M. Ridker[65,66,178], Manuel A. Rivas[179], Neil R. Robertson[3,129], Antonietta Robino[180], Igor Rudan[171], Katherine S. Ruth[181], Danish Saleheen[174,182], Veikko Salomaa[123], Nilesh J. Samani[156,157], Pamela J. Schreiner[183], Matthias B. Schulze[102,158], Robert A. Scott[31], Marcelo Segura-Lepe[61],

Xueling Sim[18,184], Andrew J. Slater[185,186], Kerrin S. Small[187], Blair H. Smith[188,189], Jennifer A. Smith[44], Lorraine Southam[3,72], Timothy D. Spector[187], Elizabeth K. Speliotes[124,125,126], Kari Stefansson[190,191], Valgerdur Steinthorsdottir[190], Kathleen E. Stirrups[16,33], Konstantin Strauch[109,192], Heather M. Stringham[18], Michael Stumvoll[45,46], Liang Sun[112], Praveen Surendran[55], Karin M. A. Swart[193], Jean-Claude Tardif[9,86], Kent D. Taylor[14], Alexander Teumer[194], Deborah J. Thompson[80], Gudmar Thorleifsson[190], Unnur Thorsteinsdottir[190,191], Betina H. Thuesen[122], Anke Tönjes[195], Mina Torres[196], Emmanouil Tsafantakis[197], Jaakko Tuomilehto[123,198,199,200], André G. Uitterlinden[20,21], Matti Uusitupa[201], Cornelia M. van Duijn[20], Mauno Vanhala[202,203], Rohit Varma[196], Sita H. Vermeulen[204], Henrik Vestergaard[50,205], Veronique Vitart[108], Thomas F. Vogt[206], Dragana Vuckovic[99,100], Lynne E. Wagenknecht[207], Mark Walker[208], Lars Wallentin[209], Feijie Wang[112], Carol A. Wang[166,167], Shuai Wang[153], Nicholas J. Wareham[31], Helen R. Warren[16,57], Dawn M. Waterworth[210], Jennifer Wessel[211], Harvey D. White[212], Cristen J. Willer[124,125,213], James G. Wilson[214], Andrew R. Wood[181], Ying Wu[175], Hanieh Yaghootkar[181], Jie Yao[14], Laura M. Yerges-Armstrong[165,215], Robin Young[55,216], Eleftheria Zeggini[72], Xiaowei Zhan[217], Weihua Zhang[60,61], Jing Hua Zhao[31], Wei Zhao[182], He Zheng[112], Wei Zhou[124,125], M. Carola Zillikens[20,21], CHD Exome+ Consortium[218], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[218], EPIC-CVD Consortium[218], ExomeBP Consortium[218], Global Lipids Genetic Consortium[218], GoT2D Genes Consortium[218], InterAct[218], ReproGen Consortium[218], T2D-Genes Consortium[218], The MAGIC Investigators[218], Fernando Rivadeneira[20,21], Ingrid B. Borecki[28], J. Andrew Pospisilik[15], Panos Deloukas[16,219], Timothy M. Frayling[181], Guillaume Lettre[9,86], Karen L. Mohlke[175], Jerome I. Rotter[14], Zoltán Kutalik[43,220], Joel N. Hirschhorn[11,13,221], L. Adrienne Cupples[27,153,226], Ruth J. F. Loos[7,8,222,226], Kari E. North[223,226*] and Cecilia M. Lindgren[3,224,226*]

[1]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA. [2]Weis Center for Research, Geisinger Health System, Danville, PA, USA. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, Cyprus. [5]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. [6]Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [7]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Montreal Heart Institute, Universite de Montreal, Montreal, Quebec, Canada. [10]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. [11]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [12]Department of Genetics, Harvard Medical School, Boston, MA, USA. [13]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [14]Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA. [15]Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [16]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [17]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [18]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [19]McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, USA. [20]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [21]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [22]Estonian Genome Center, University of Tartu, Tartu, Estonia. [23]Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA. [24]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [25]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [26]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [27]NHLBI Framingham Heart Study, Framingham, MA, USA. [28]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [29]Department of Medicine, Harvard University Medical School, Boston, MA, USA. [30]Massachusetts General Hospital, Boston, MA, USA. [31]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK. [32]Department of Vascular Medicine, AMC, Amsterdam, The Netherlands. [33]Department of Haematology, University of Cambridge, Cambridge, UK. [34]School of Public Health, University of Michigan, Ann Arbor, MI, USA. [35]School of Kinesiology and Health Science, Faculty of Health, York University, Toronto, Canada. [36]Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA, USA. [37]INSERM U1167, Lille, France. [38]Institut Pasteur de Lille, U1167, Lille, France. [39]U1167—RID-AGE, Universite de Lille - Risk factors and molecular determinants of aging-related diseases, Lille, France. [40]INSERM U1018, Centre de recherche en Épidemiologie et Sante des Populations (CESP), Villejuif, France. [41]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA. [42]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [43]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [44]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [45]IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany. [46]Department of Medicine, University of Leipzig, Leipzig, Germany. [47]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, Germany. [48]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [49]Department of Nephrology, University Hospital Regensburg, Regensburg,

Germany. [50]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [51]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [52]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark. [53]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [54]Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA. [55]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [56]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [57]NIHR Barts Cardiovascular Research Centre, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, London, UK. [58]Research Centre on Public Health, University of Milano-Bicocca, Monza, Italy. [59]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [60]Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex, UK. [61]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [62]Imperial College Healthcare NHS Trust, London, UK. [63]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [64]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [65]Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA, USA. [66]Harvard Medical School, Boston, MA, USA. [67]Medical department, Lillebaelt Hospital, Vejle, Denmark. [68]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [69]Department of Biostatistics, University of Liverpool, Liverpool, UK. [70]Menzies Health Institute Queensland, Griffith University, Southport, Queensland, Australia. [71]Department of Biomedical Infomatics and Medical Education, University of Washington, Seattle, WA, USA. [72]Wellcome Trust Sanger Institute, Hinxton, UK. [73]British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK. [74]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. [75]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. [76]Faculty of Pharmacy, Universite de Montreal, Montreal, Quebec, Canada. [77]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. [78]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [79]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [80]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [81]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [82]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. [83]Institute of Cardiovascular Science, University College London, London, UK. [84]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. [85]Department of Life Sciences, Brunel University London, Uxbridge, UK. [86]Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, Canada. [87]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [88]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [89]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. [90]CNR Institute of Clinical Physiology, Pisa, Italy. [91]Department of Clinical & Experimental Medicine, University of Pisa, Pisa, Italy. [92]Toulouse University School of Medicine, Toulouse, France. [93]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. [94]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmo, Sweden. [95]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA. [96]Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, Sweden. [97]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [98]Ilaria Gandin, Research Unit, AREA Science Park, Trieste, Italy. [99]Department of Medical Sciences, University of Trieste, Trieste, Italy. [100]Institute for Maternal and Child Health-IRCCS Burlo Garofolo, Trieste, Italy. [101]Geriatrics, Department of Public Health, Uppsala University, Uppsala, Sweden. [102]German Center for Diabetes Research, München-Neuherberg, Germany. [103]Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [104]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [105]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [106]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, USA. [107]University of Exeter Medical School, University of Exeter, Exeter, UK. [108]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [109]Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [110]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. [111]Department of Vascular Medicine, AMC, Amsterdam, The Netherlands. [112]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [113]Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. [114]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [115]HUNT Research Center, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway. [116]Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands. [117]Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands. [118]Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. [119]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [120]Faculty of Medicine, Aalborg University, Aalborg, Denmark. [121]Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [122]Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, Denmark. [123]National Institute for Health and Welfare, Helsinki, Finland. [124]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [125]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [126]Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA. [127]Centre for Brain Research, Indian Institute of Science, Bangalore, India. [128]Echinos Medical Centre, Echinos, Greece. [129]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [130]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. [131]UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK. [132]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [133]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London, UK. [134]University Medical Centre Mannheim, 5th Medical Department, University of Heidelberg, Mannheim, Germany. [135]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [136]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, Finland. [137]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [138]Verge Genomics, San Fransico, CA, USA. [139]Division of Biomedical and Personalized Medicine, Department of Medicine, University of Colorado-Denver, Aurora, CO, USA. [140]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [141]Department of Health Services, University of Washington, Seattle, WA, USA. [142]Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, Philippines. [143]USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines. [144]Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. [145]Department of Social Work, Tunghai University, Taichung, Taiwan. [146]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [147]Department of Clinical Chemistry, Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [148]Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [149]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Palo Alto, CA, USA. [150]Uppsala University, Uppsala, Sweden. [151]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark. [152]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [153]Department of

Biostatistics, Boston University School of Public Health, Boston, MA, USA. [154]Department of Public Health Sciences, Institute for Personalized Medicine, The Pennsylvania State University College of Medicine, Hershey, PA, USA. [155]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [156]Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester, UK. [157]NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [158]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, Germany. [159]Department of Medicine, Kuopio University Hospital, Kuopio, Finland. [160]Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, France. [161]Department of Public Health, University Hospital of Strasbourg, Strasbourg, France. [162]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands. [163]Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitat, Munich, Germany. [164]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. [165]Program for Personalized and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [166]Division of Obstetric and Gynaecology, School of Medicine, The University of Western Australia, Perth, Western Australia, Australia. [167]School of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Newcastle, New South Wales, Australia. [168]Institute for Molecular Medicine (FIMM) and Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland. [169]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [170]School of Medicine, University of Split, Split, Croatia. [171]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [172]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [173]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [174]Centre for Non-Communicable Diseases, Karachi, Pakistan. [175]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [176]Genetics and Pharmacogenomics, Merck & Co., Inc., Boston, MA, USA. [177]Department of Epidemiology, University of Washington, Seattle, WA, USA. [178]Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [179]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [180]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [181]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [182]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [183]Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN, USA. [184]Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore. [185]Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, USA. [186]OmicSoft a QIAGEN Company, Cary, NC, USA. [187]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [188]Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [189]Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK. [190]deCODE Genetics/Amgen Inc., Reykjavik, Iceland. [191]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [192]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany. [193]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands. [194]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [195]Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig, Leipzig, Germany. [196]USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. [197]Anogia Medical Centre, Anogia, Greece. [198]Centre for Vascular Prevention, Danube-University Krems, Krems, Austria. [199]Dasman Diabetes Institute, Dasman, Kuwait. [200]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [201]Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [202]Central Finland Central Hospital, Jyvaskyla, Finland. [203]University of Eastern Finland, Kuopio, Finland. [204]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. [205]Steno Diabetes Center Copenhagen, Gentofte, Denmark. [206]Cardiometabolic Disease, Merck & Co., Inc., Kenilworth, NJ, USA. [207]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [208]Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, UK. [209]Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. [210]Genetics, Target Sciences, GlaxoSmithKline, Collegeville, PA, USA. [211]Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks School of Public Health & School of Medicine, Indiana University, Indiana, IN, USA. [212]Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand. [213]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [214]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. [215]GlaxoSmithKline, King of Prussia, PA, USA. [216]University of Glasgow, Glasgow, UK. [217]Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA. [218] A list of members and affiliations appears in the Supplementary Note. [219]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [220]Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland. [221]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA. [222]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [223]Department of Epidemiology and Carolina Center of Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [224]Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, UK. [225]These authors contributed equally: Anne E. Justice, Tugce Karaderi, Heather M. Highland, Kristin L. Young, Mariaelisa Graff, Yingchang Lu, Valérie Turcot. [226]These authors jointly supervised this work: L. Adrienne Cupples, Ruth J.F. Loos, Kari E. North and Cecilia M. Lindgren. *e-mail: kari_north@unc.edu; celi@well.ox.ac.uk

Methods

Studies. Stage 1 included 74 studies (12 case/control, 59 population-based and five family) comprising 344,369 adults of European ($N=288,492$), African ($N=15,687$), South Asian ($N=29,315$), East Asian ($N=6,800$) and Hispanic ($N=4,075$) descent. Stage 1 meta-analyses were conducted in each ancestry and in all ancestries together, for both sex-combined and sex-specific analyses. Follow-up analyses were performed in 132,177 individuals of European ancestry from deCODE and the UK Biobank, Release 1[112] (UKBB) (Supplementary Data 1–3). Informed consent was obtained by the parent study and protocols approved by each study's institutional review boards.

Phenotypes. For each study, WHR (waist circumference divided by hip circumference) was corrected for age, BMI and genomic principal components (derived from GWAS data, the variants with MAF >1% on the ExomeChip, and ancestry informative markers available on the ExomeChip), as well as any additional study-specific covariates (for example, recruiting center), in a linear regression model. For studies with unrelated individuals, residuals were calculated separately by sex, whereas for family based studies sex was included as a covariate in models with both men and women. Residuals for case/control studies were calculated separately. Finally, residuals were inverse normal transformed and used as the outcome in association analyses. Phenotype descriptives by study are shown in Supplementary Data 3.

Genotypes and quality control. Most studies followed a standardized protocol and performed genotype calling using the algorithms indicated in Supplementary Data 2, which typically included zCall[3]. For 10 studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, raw intensity data for samples from seven genotyping centers were combined for joint calling[4]. Study-specific quality control of the genotyped variants was implemented before association analysis (Supplementary Data 1,2). To assess whether any significant associations with rare and low-frequency variants could be due to allele calling in smaller studies, we performed a sensitivity meta-analysis of all large studies (>5,000 participants) compared to all studies. We found very high concordance for effect sizes, suggesting that smaller studies do not bias our results (Supplementary Fig. 24).

Study-level statistical analyses. Each cohort performed single-variant analyses for both additive and recessive models in each ancestry, for sexes combined and sex-specific groups, with either RAREMETALWORKER (see URLs) or RVTESTs (see URLs). to associate inverse normal transformed WHRadjBMI with genotype accounting for cryptic relatedness (kinship matrix) in a linear mixed model. Both programs perform score-statistic rare-variant association analysis, accommodate unrelated and related individuals and provide single-variant results and variance-covariance matrices. The covariance matrix captures linkage disequilibrium between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[113,114].

Centralized quality control. Individual cohorts identified ancestry outliers based on 1000 Genomes Phase 1 reference populations. A centralized quality control procedure implemented in EasyQC[115] was applied to individual cohort summary statistics to identify cohort-specific problems: (1) possible errors in phenotype residual transformation; (2) strand issues and (3) inflation due to population stratification, cryptic relatedness and genotype biases.

Meta-analyses. Meta-analyses were carried out in parallel by two analysts at two sites using RAREMETAL[113]. We excluded variants if they had call rate <95%, Hardy–Weinberg equilibrium $P<1\times10^{-7}$, or large allele frequency deviations from reference populations (>0.6 for all ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded markers not present on the Illumina ExomeChip array 1.0, Y-chromosome and mitochondrial variants, indels, multiallelic markers and problematic variants on the basis of Blat-based sequence alignment. Significance for single-variant analyses was defined at an array-wide level ($P<2\times10^{-7}$). For all suggestive significant variants ($P<2\times10^{-6}$) from stage 1, we calculated P_{sexhet} for each SNP, testing for differences between women-specific and men-specific beta estimates and standard errors using EasyStrata[116]. Each SNP that reached $P_{sexhet}<0.05$ per number of variants tested (70 variants brought forward from stage 1, $P_{sexhet}<7.14\times10^{-4}$) was considered significant. Additionally, while each individual study was asked to perform association analyses stratified by ancestry and adjusted for population stratification, all study-specific summary statistics were combined in our all ancestry meta-analyses. To investigate potential heterogeneity across ancestries, we examined ancestry-specific meta-analysis results for our top 70 variants from stage 1 and found no evidence of significant across-ancestry heterogeneity for any of our top variants (I^2 values noted in Supplementary Data 4–6).

For the gene-based analyses, we applied two sets of criteria to select variants with a MAF <5% in each ancestry based on coding variant annotation from five prediction algorithms (PolyPhen2, HumDiv and HumVar, LRT, MutationTaster and SIFT)[117]. Our broad gene-based tests included nonsense, stop-loss, splice site and missense variants annotated as damaging by at least one algorithm mentioned above. Our strict gene-based tests included only nonsense, stop-loss, splice site and missense variants annotated as damaging by all five algorithms. These analyses were performed using the sequence kernel association test (SKAT) and variable threshold methods in RAREMETAL[113]. Statistical significance for gene-based tests was set at a Bonferroni-corrected threshold of $P<2.5\times10^{-6}$ (0.05/~20,000 genes).

Genomic inflation. We observed marked genomic inflation of the test statistics even after controlling for population stratification arising mainly from common markers; λ_{GC} in the primary meta-analysis (combined ancestries and combined sexes) was 1.06 for all variants and 1.37 for common coding and splice site markers, respectively (Supplementary Figs. 3, 7 and 13 and Supplementary Table 10). Such inflation is expected for a highly polygenic trait such as WHRadjBMI, for studies using a non-random set of variants across the genome, and is consistent with our very large sample size[115,118,119].

Conditional analyses. The RAREMETAL R-package[113] was used to identify independent WHRadjBMI association signals across all ancestries and European meta-analysis results. RAREMETAL performs conditional analyses using covariance matrices to distinguish true signals from shadows of adjacent significant variants in linkage disequilibrium. First, we identified lead variants ($P<2\times10^{-7}$) based on a 1 Mb window centered on the most significant variant. We then conditioned on the lead variants in RAREMETAL and kept new lead signals at $P<2\times10^{-7}$ for conditioning in a second round of analysis. The process was repeated until no additional signal emerged below the pre-specified P value threshold ($P<2\times10^{-7}$).

To test if the associations detected were independent of previously published WHRadjBMI variants[10,14,16], we used RAREMETAL to perform conditional analyses in the stage 1 discovery set if the GWAS variant or its proxy ($r^2\geq0.8$) was on the ExomeChip. All variants identified in our meta-analysis and the previously published variants were available in the UKBB dataset[112], which was used as a replacement dataset if a good proxy was not on the ExomeChip. All conditional analyses in the UKBB were performed using SNPTEST[120–122].

The conditional analyses were carried out reciprocally, conditioning on the ExomeChip variant and then the previously published variant. An association was considered independent if it was significant before conditional analysis ($P<2\times10^{-7}$) with both the exome chip variant and the previously published variant, and the observed association with our variant remained significant on conditional analysis. Conditional P values between 9×10^{-6} and 0.05 were considered inconclusive, while those $<9\times10^{-6}$ were considered suggestive.

Stage 2 meta-analyses. In stage 2, we sought to validate 70 stage 1 variants ($P<2\times10^{-6}$) in two independent studies, UKBB ($N=119,572$) and deCODE ($N=12,605$), using the same quality control and analytical methodology. Genotyping, study descriptions and phenotype descriptives are provided in Supplementary Data 1–3. Stage 1 and 2 meta-analysis was performed using the inverse-variance weighted fixed-effects method. Significant associations were defined as those nominally significant ($P<0.05$) in stage 2 when available in stage 2, and array-wide significance for stage 1 and 2 at $P<2\times10^{-7}$ (0.05/~250,000 246,328 variants tested). Variants are considered novel and statistically significant if they were greater than 1 Mb from a previously identified WHRadjBMI lead SNP[10–16] and achieved a significance threshold of $P<2\times10^{-7}$.

Pathway enrichment analyses: EC-DEPICT. We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the ExomeChip ('EC-DEPICT') described further in a companion manuscript[21]. DEPICT uses 'reconstituted' gene sets, where different types of gene set (for example, canonical pathways, protein-protein interaction networks, and mouse phenotypes) were extended through large-scale microarray data (see Pers et al.[20] for details). EC-DEPICT computes P values based on Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS) and Scania Diabetes Registry (SDR) cohorts, $N=11,899$) and, unlike DEPICT, takes as input only genes directly containing substantial (coding) variants rather than all genes in a specified linkage disequilibrium (Supplementary Note).

Two analyses were performed for WHRadjBMI ExomeChip: one with all variants $P<5\times10^{-4}$ (49 significant gene sets in 25 meta-gene sets, FDR < 0.05) and one with all variants >1 Mb from known GWAS loci[10] (26 significant gene sets in 13 meta-gene sets, FDR < 0.05). Affinity propagation clustering[123] was used to group highly correlated gene sets into 'meta-gene sets'; for each meta-gene set, the member gene set with the best P value was used for visualization (Supplementary Note). EC-DEPICT was written in Python (see URLs).

Pathway enrichment analyses: PASCAL. We also applied PASCAL pathway analysis[22] to summary statistics from stage 1 for all coding variants. PASCAL derives gene-based scores (SUM and MAX) and tests for over-representation of high gene scores in predefined biological pathways. We performed both MAX and SUM estimations for pathway enrichment. MAX is sensitive to gene sets driven by a single signal, while SUM is better for multiple variant associations in the same gene. We used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and also added dichotomized ($z>3$) reconstituted gene sets

from DEPICT[20]. To accurately estimate SNP-by-SNP correlations even for rare variants, we used the UK10K data (TwinsUK[124] and ALSPAC[125], $N=3781$). To distinguish contributions of regulatory and coding variants, we also applied PASCAL to summary statistics of only regulatory variants (20 kb upstream) and regulatory+coding variants from the Shungin et al.[10] study. In this way, we could investigate what is gained by analyzing coding variants.

Monogenic obesity enrichment analyses. We compiled two lists consisting of 31 genes with strong evidence that disruption causes monogenic forms of insulin resistance or diabetes, and eight genes with evidence that disruption causes monogenic forms of lipodystrophy. To test for association enrichment, we conducted simulations by matching each gene with others based on gene length and number of variants tested to create 1,000 matched gene sets and assessed how often the number of variants exceeding set significance thresholds was greater than in our monogenic obesity gene set.

Variance explained. We estimated phenotypic variance explained by stage 1 associations in all ancestries for men, women, and combined sexes[126]. For each associated region, we pruned subsets of SNPs within 500 kb of SNPs with the lowest P value and used varying P value thresholds (ranging from 2×10^{-7} to 0.02) from the combined sexes results. Additionally, we examined all variants and independent variants across a range of MAFs. The variance explained by each subset of SNPs in each stratum was estimated by summing the variance explained by individual top coding variants. To compare variance explained between men and women, we tested for significant differences assuming the weighted sum of χ^2-distributed variables tend to a Gaussian distribution following Lyapunov's central limit theorem[126,127].

Cross-trait lookups. To evaluate relationships between WHRadjBMI and related cardiometabolic, anthropometric, and reproductive traits, association results for the 51 WHRadjBMI coding SNPs were requested from seven consortia, including ExomeChip data from GIANT (BMI, height), Global Lipids Genetics Consortium (GLGC) (total cholesterol, triglycerides, HDL-cholesterol, LDL-cholesterol), International Consortium for Blood Pressure (IBPC)[128] (systolic and diastolic blood pressure), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) (glycemic traits) and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (type 2 diabetes))[21,25–29]. For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data released by CARDIoGRAMplusC4D[129] and for age at menarche and menopause, we used a combination of ExomeChip and 1000 Genomes Project-imputed GWAS data from ReproGen. Heat maps were generated with gplots (R v.3.3.2) using Euclidean distance based on P value, and direction of effect and complete linkage clustering (see URLs).

GWAS catalog lookups. To determine whether substantial coding variants were associated with any related cardiometabolic or anthropometric traits, we also searched the NHGRI-EBI GWAS Catalog for previous associations near our lead SNPs (+/−500 kb). We used PLINK to calculate linkage disequilibrium using ARIC European participants. All GWAS Catalog SNPs in the specified regions with an $r^2>0.7$ were evaluated[37]. Consistent direction of effect was based on the WHRadjBMI-increasing allele, linkage disequilibrium and allele frequency. We do not comment on direction of effect when a GWAS Catalog variant was not identical or in high linkage disequilibrium ($r^2>0.9$) with the WHR variant and MAF >45%.

Body-fat percentage associations. We performed body-fat percentage and truncal fat percentage lookups of 48 of the 56 WHRadjBMI identified variants (Tables 1 and 2) available in UKBB. GWAS for body-fat percentage and truncal fat percentage in UKBB excluded pregnant or possibly pregnant women, individuals with BMI <15, or those without genetically confirmed European ancestry, resulting in a sample size of 120,286. Estimated body-fat percentage and truncal fat percentage were obtained using the Tanita BC418MA body composition analyzer (Tanita). Participants were non-fasting and did not follow any specific instructions before bioimpedance measurements. SNPTEST was used to perform the analyses based on residuals adjusted for age, 15 principal components, assessment center and the genotyping chip[120].

Collider bias. To evaluate SNPs for possible collider bias[17], we used results from a recent GIANT BMI GWAS[25]. For each significant SNP from our additive models, WHRadjBMI associations were corrected for potential bias due to associations between each variant and BMI (Supplementary Note). Variants meeting Bonferroni-corrected significance ($P_{corrected}<9.09\times10^{-4}$, 0.05/55 variants examined) were considered robust against collider bias.

***Drosophila* RNAi-knockdown experiments.** For each gene with WHRadjBMI-associated coding variants in the final combined meta-analysis ($P<2\times10^{-7}$), its corresponding *Drosophila* orthologs were identified in the Ensembl ortholog database (see URLs), when available. *Drosophila* triglyceride content values were mined from a publicly available genome-wide fat screen dataset[45] to identify potential genes for follow-up knockdowns. Estimated values represent fractional changes in triglyceride content in adult male flies. Data are from male progeny of crosses between male UAS-RNAi flies from the Vienna Drosophila Resource Center and Hsp70-GAL4; Tub-GAL8ts virgin females. (Supplementary Note). The screen comprised one to three biological replicates. We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride content.

Orthologs for two genes were brought forward for follow-up, *DNAH10* and *PLXND1*. For both genes, we generated adipose tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses to knockdown transcripts in a tissue-specific manner, leveraging upstream activation sequence (UAS)-inducible short-hairpin knockdown lines, available through the Vienna Drosophila Resource Center. Specifically, elav-Gal4, which drives expression of the RNAi construct in post mitotic neurons starting at embryonic stages all the way to adulthood, was used. Cg drives expression in the fat body and hemocytes starting at embryonic stage 12, all the way to adulthood. (Supplementary Note). Resulting triglyceride values were normalized to fly weight and larval/population density. We used the non-parametric Kruskall–Wallis test to compare wild type with knockdown lines.

eQTLs analysis. We queried the significant variant (Exome coding SNPs)-gene pairs associated with eGenes across five metabolically relevant tissues (skeletal muscle, subcutaneous adipose, visceral adipose, liver and pancreas) with at least 70 samples in the GTEx database[46]. For each tissue, variants were selected based on the following thresholds: the minor allele was observed in at least 10 samples and MAF ≥1%. eGenes, genes with a significant eQTL, are defined on a FDR[130] threshold of ≤0.05 of beta distribution-adjusted empirical P value from FastQTL. Nominal P values were generated for each variant-gene pair by testing the alternative hypothesis that the slope of a linear regression model between genotype and expression deviates from zero. To identify all significant variant-gene pairs associated with eGenes, a genome-wide empirical P value threshold[64] (pt) was defined as the empirical P value of the gene closest to the 0.05 FDR threshold. pt was then used to calculate a nominal P value threshold for each gene based on the beta distribution model (from FastQTL) of the minimum P value distribution f(pmin) obtained from the permutations for the gene. For each gene, variants with a nominal P value below the gene-level threshold were considered significant and included in the final list of variant-gene pairs[64]. For each eGene, we also listed the most significantly associated variants (eSNP). Only these exome SNPs with $r^2>0.8$ with eSNPs were considered for biological interpretation (supplementary eQTL GTEx).

We also performed cis-eQTL analysis in 770 METSIM subcutaneous adipose tissue samples as described in Civelek et al.[131]. A FDR was calculated using all P values from the cis-eQTL detection in the q-value package in R. Variants associated with nearby genes at an FDR of less than 1% were considered to be significant (equivalent $P<2.46\times10^{-4}$).

For loci with more than one microarray probeset of the same gene associated with the exome variant, we selected the probeset that provided the strongest linkage disequilibrium r^2 between the exome variant and the eSNP. In reciprocal conditional analysis, we conditioned on the lead exome variant by including it as a covariate in the cis-eQTL detection and reporting the P value of the eSNP and vice versa. Signals were considered coincident if both the lead exome variant and the eSNP were no longer significant after conditioning on the other and the variants were in high linkage disequilibrium ($r^2>0.80$).

For loci that also harbored reported GWAS variants, we performed reciprocal conditional analysis between the GWAS lead variant and the lead eSNP. For loci with more than one reported GWAS variant, the GWAS variant with the strongest linkage disequilibrium I^2 with the lead eSNP was reported.

Penetrance analysis. Phenotype and genotype data from UKBB were used for penetrance analysis. Three of 16 rare and low-frequency variants (MAF ≤1%) detected in the final stage 1 and 2 meta-analysis were available in UKBB and had relatively larger effect sizes (>0.90). Phenotype data for these three variants were stratified by WHR using World Health Organization (WHO) guidelines, which consider women and men with WHR greater than 0.85 and 0.90 as obese, respectively. Genotype and allele counts were used to calculate the number of carriers of the minor allele. The number of obese versus non-obese carriers for women, men and sexes combined was compared using a χ^2 test. Significance was determined using a Bonferroni correction for the number of tests performed ($0.05/9=5.5\times10^{-3}$).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Summary statistics of all analyses are available at https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

References

112. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

113. Liu, D. J. et al. Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
114. Goldstein, J. I. et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).
115. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
116. Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
117. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
118. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
119. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, S1–S3 (2012).
120. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
121. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
122. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
123. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
124. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* **42**, 76–85 (2013).
125. Boyd, A. et al. Cohort profile: the 'children of the 90s'–the index offspring of the Avon Longitudinal Study of parents and children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
126. Kutalik, Z., Whittaker, J., Waterworth, D., Beckmann, J. S. & Bergmann, S. Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. *Genet. Epidemiol.* **35**, 341–349 (2011).
127. Billingsley, P. *Probability and Measure* (Wiley, New York, 1986).
128. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
129. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
130. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
131. Civelek, M. et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).

nature research

Corresponding author(s): NG-A47611R

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	None
Data analysis	RareMetalWorker version 4.13; RVtests version 20140416; rareMETALs R-package version 5.8; EasyQC version 8.6; EasyStrata version 8.6; Eigensoft version 3.0; PLINK version 1.9; R version 3.0.3 SNPRelate version 0.9.19

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size in association and expression quantitative trait loci studies were determined from individuals that pass quality control for genotyping and have phenotypes available. For the Drosophila experiment the flies were manually handled and counted and all experiments were performed with one to three biological replicates. We used Quanto version 1.2.4 to determine power given the sample size and allele frequencies. See Figure 2.
Data exclusions	Association studies Individual cohorts identified ancestry population outliers based on 1000 Genome Project phase 1 ancestry reference populations. A centralized quality-control procedure implemented in EasyQC was applied to individual cohort association summary statistics to identify cohort-specific problems: (1) assessment of possible errors in phenotype residual transformation; (2) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues, and (3) examination of quantile-quantile (QQ) plots per study to identify any inflation arising from population stratification, cryptic relatedness and genotype biases. Meta-analyses During the meta-analyses, we excluded variants if they had call rate <95%, Hardy-Weinberg equilibrium P-value <1x10-7, or large allele frequency deviations from reference populations (>0.6 for all ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded from downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y-chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based on the Blat-based sequence alignment analyses. Study-specific quality control (QC) measures of the genotyped variants were implemented before association analysis (Supplementary Tables 1-2) (this is far to extensive to outline here) - all exclusion criteria and protocols have been implemented and used previously and extensively published on.
Replication	Association studies We validated a total of 70 variants from Stage 1 that met P<2x10-6 in two 1219 independent studies, the UK Biobank (Release 1) and Iceland (deCODE), comprising 119,572 and 12,605 individuals, respectively. The same QC and analytical methodology were used for these studies. For the combined analysis of Stage 1 plus 2, we used the inverse-variance weighted fixed effects meta-analysis method. Significant associations were defined as those nominally significant (P<0.05) in the Stage 2 study and for the combined meta-analysis (Stage 1 plus Stage 2) significance was set at P<2x10-7 (0.05/~250,000 variants). Variants that were not available in Stage 2 studies are noted in text and tables. Drosophila studies Flies were investigated pooled in groups of eight, and the screen comprised one to three biological replicates. We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride content. Thus, orthologues for two genes were brought forward for follow-up, DNAH10 and PLXND1. For both genes, we generated adipose tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses to knockdown transcripts in a tissue specific manner, leveraging upstream activation sequence (UAS)- inducible short-hairpin knockdown lines, available through the VDRC (Vienna Drosophila Resource Center) to confirm the initial observations. Five-to-seven-day-old males were sorted into groups of 20, and the experiment was carried out in duplicate according to protocol, with one alteration: the samples were cleared of residual particulate debris by centrifugation before absorbance reading. Resulting triglyceride values were normalized to fly weight and larval/ population density. We only highlight the results that are reliably reproduced.
Randomization	In the quantitative trait analysis this is not relevant as there is no allocation into groups.
Blinding	NA

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Data are from male progeny of crosses between male UAS-RNAi flies from the Vienna Drosophila Resource Center (VDRC) and Hsp70-GAL4; Tub-GAL8ts virgin females. (Additional details are found within the Online Methods and Supplementary Note).
Wild animals	NA
Field-collected samples	NA